DOCUMENT RESUME

ED 334 241 TM 016 774

AUTHOR Rock, Donald A.; Pollack, Judith M.
TITLE Psychometric Report for the NELS:88 Base Year Test Battery. Contractor Report. Technical Report.
INSTITUTION Educational Testing Service, Princeton, N.J.
SPONS AGENCY National Center for Education Statistics (ED), Washington, DC.
REPORT NO NCES-91-468
PUB DATE Apr 91
CONTRACT 300-86-0010
NOTE 150p.; Data Series: NELS:88-88-1.4.
PUB TYPE Statistical Data (110) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC06 Plus Postage.
DESCRIPTORS *Achievement Tests; Citizenship Education; Geography; *Grade 8; History; Junior High School Students; *Longitudinal Studies; Mathematics Tests; *National Surveys; Postsecondary Education; Psychometrics; Reading Comprehension; Reading Tests; Science Tests; Secondary Education; Statistical Data; Tables (Data); *Test Results
IDENTIFIERS *Baseline Data; *National Education Longitudinal Study 1988

ABSTRACT

The National Education Longitudinal Study of 1988 (NELS:88) is designed to monitor the transition of a national sample of young adults as they progress from junior to senior high school and then to postsecondary education or the world of work. An in-depth description is provided of the rationale, development, and psychometric properties of the base year test for grade 8. The achievement test battery was composed of four tests; (1) reading comprehension; (2) mathematics; (3) science; and (4) history/citizenship/geography. The eighth grade (base year) sample was composed of approximately 24,600 eighth graders from 1,052 schools. Results show that the NELS:88 test battery met or exceeded all its psychometric objectives. Reliabilities for the reading comprehension, mathematics, and history/citizenship/geography tests were acceptable; the science test was somewhat less reliable. Internal consistency was high enough to justify item response theory scoring. There was no consistent evidence of item bias for gender or racial/ethnic groups. Factor analyses support the discriminant validity of the four content areas tested. Five tables and seven figures complement the discussion. A 32-item list of references is included. Eight appendices provide item analysis statistics, differential item functioning statistics, item parameters, test information functions, descriptions of individual items, intercorrelations of testlets, definitions of proficiency scores, and standard errors of measurement at theta scale points. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

NATIONAL CENTER FOR EDUCATION STATISTICS

Technical Report April 1991

Psychometric Report for the NELS:88 Base Year Test Battery

Contractor Report

Donald A. Rock
Judith M. Pollack
Educational Testing Service

Jeffrey Owings
Anne Hafner
Project Officers
National Center for Education Statistics

Data Series:
NELS:88–88–1.4

U.S. Department of Education
Office of Educational Research and Improvement NCES 91-468

U.S. Department of Education
Lamar Alexander
Secretary

Office of Educational Research and Improvement
Christopher T. Cross
Assistant Secretary

National Center for Education Statistics
Emerson J. Elliott
Acting Commissioner

National Center for Education Statistics

"The purpose of the Center shall be to collect, and analyze, and disseminate statistics and other data related to education in the United States and in other nations."—Section 406(b) of the General Education Provisions Act, as amended (20 U.S.C. 1221e–1).

April 1991

Contact:
Jeffrey Owings
(202) 219–1777

Prepared for the U.S. Department of Education, National Center for Education Statistics, under contract No. 300–86–0010. Contractors undertaking such projects are encouraged to freely express their professional judgment. This report, therefore, does not necessarily represent positions or policies of the federal government, and no official endorsement should be inferred. This report is released as received from the contractor.

EXECUTIVE SUMMARY

The National Education Longitudinal Study of 1988 (NELS:88) is sponsored by the National Center for Education Statistics (NCES) and is designed to monitor the transition of a national sample of young adults as they progress from junior to senior high school and then on to postsecondary education and/or the world of work. The primary purpose of the NELS:88 longitudinal study is to provide policy-relevant information on the effectiveness of schools, curriculum paths, special programs, variations in curriculum content, and/or mode of delivery in bringing about educational growth.

Among the more important educational indicators that will be monitored at the eighth, tenth, and twelfth grade is the achievement test battery. The NELS:88 test battery is composed of four separate tests--Reading Comprehension, Mathematics, Science, and History/Citizenship/Geography. The NELS:88 test battery is critical to the measurement of growth in educational achievement that will take place during the last four years of secondary schooling. In addition to providing trend information on academic achievement for its longitudinal cohort, the test battery is also designed to provide cross-sectional trend information when comparisons are made with the 1980 High School and Beyond cohorts.

The NELS:88 base year (eighth grade) sample was composed of approximately 24,600 eighth graders who were sampled from 1,052 schools.

This report provides an in-depth description of the rationale, development, and psychometric properties of the eighth grade test.

The results suggest that the NELS:88 test battery either met or exceeded all of its psychometric objectives. The eighth grade analysis indicated that:

- While the allotted testing time was only one and a half hours, quite acceptable reliabilities were obtained for the Reading Comprehension, Mathematics, History/Citizenship/Geography, and to a somewhat lesser extent the Science test.

- The internal consistency reliabilities were sufficiently high to justify the use of Item Response Theory (IRT) scoring, and thus provide the framework for constructing tenth and twelfth grade forms that will be adaptive to the ability level of the student. The IRT scaling will enable the researcher to administer forms varying in difficulty at the tenth grade and to scale these scores on a common metric. The choice of test form administered to a student in grade ten will be determined by the relative ability level demonstrated by the student in grade eight. This adaptive approach will both minimize potential ceiling effects and increase measurement accuracy when the students are followed up in the tenth and twelfth grades.

- There was no consistent evidence of differential item functioning (item bias) for either gender or racial/ethnic groups.

- Factor analytic results supported the discriminant validity of the four tested content areas. Convergent validity was also indicated by salient loadings of testlets composed of "marker items" on their hypothesized factors.

- In addition to providing the usual normative scores in all four tested areas, behaviorally anchored proficiency scores have been provided in both the Reading and Mathematics areas.

ACKNOWLEDGMENTS

We would like to especially thank the National Center for Education Statistics (NCES) staff members Anne Hafner and Jeffrey Owings and the other reviewers for their reviews and helpful comments with respect to the final revision of this report. Reviewers included Gary Phillips of NCES, John Gugel of NCES, Larry Ogle of NCES, Steve Gorman of NCES and Leigh Burstein from the University of California in Los Angeles. We would also like to thank the Technical Review Panel assembled by the NCES staff whose suggestions in the test development stage were particularly helpful in developing the final test specifications for the NELS:88 8th grade test battery. Also Sindy McGill and Ruth Harris of NCES for their valuable assistance during the production of the final report.

CONTENTS

EXECUTIVE SUMMARY iii
ACKNOWLEDGMENTS v
LIST OF TABLES viii
LIST OF FIGURES ix

CHAPTER 1. INTRODUCTION 1

CHAPTER 2. TEST SPECIFICATIONS 3

Aims and Objectives 3
Two Stage Testing in a Longitudinal Framework 4
Difficulty Level 5
IRT Scaling for Longitudinal Measurement and Equating to Earlier Cohorts 6
Psychometric Goals of the NELS:88 Eighth Grade Test Battery 6
Specifications for Individual Tests 8
Reading 8
Mathematics 11
Science 13
History/Citizenship/Geography 14
Matching Test Content to Curriculum 15

CHAPTER 3. PSYCHOMETRIC ANALYSIS RESULTS 19

Were the Tests Speeded? 19
Reliabilities of the NELS:88 Eighth Grade Test Battery 19
Item Statistics by Gender and Racial/Ethnic Groups 22
Differential Item Functioning (DIF) 25
Factor Structure of the NELS:88 Eighth Grade Test Battery 28
Performance of Racial/Ethnic and Gender Groups on the NELS:88 Eighth Grade Test Battery 31
Proficiency Level Subscores by Subgroups 31
Item Response Theory (IRT) Parameters for the NELS:88 Battery 37
Test Scores on User Tape 42

CHAPTER 4. CONCLUSIONS 45

REFERENCES 47
APPENDICES 51

LIST OF TABLES

Table

1 Speededness Indices for Tests by Racial/Ethnic and Sex Groups (Percent of Sample Who Reached Last Item) 19

2 Test Reliabilities and Standard Errors of Measurement by Race/Ethnicity and Sex 21

3 A Comparison of Average Difficulty and Average Biserials for Comparable Tests in the HS&B and NELS:88 Test Battery 24

4 Factor Structure, NELS:88 Tests 30

5 Weighted Means and Standard Deviations of IRT Scores on the NELS:88 Tests by Racial/Ethnic Groups and Sex 32

LIST OF FIGURES

Figure

1 Reading Test Specifications (Number of Items by Process and Content) 10

2 Mathematics Specifications (Number of Items by Process and Content) 12

3 Science Test Specifications (Number of Items by Process and Content) 14

4 Percent of Selected Subgroups that are Proficient at Each Mathematics Proficiency Level 35

5 Percent of Gender Groups that are Proficient at Each Mathematics Proficiency Level 36

6 Percent of Selected Subgroups that are Proficient at Each Reading Proficiency Level 38

7 Percent of Gender Groups that are Proficient at Each Reading Proficiency Level 39

APPENDICES

APPENDIX A - ITEM ANALYSIS STATISTICS 51

A1 Reading 53
A2 Mathematics 56
A3 Science 59
A4 History/Citizenship/Geography 62

APPENDIX B - DIFFERENTIAL ITEM FUNCTIONING (DIF) 65

B1 Reading 67
B2 Math 72
B3 Science 77
B4 History/Citizenship/Geography 82

APPENDIX C - ITEM PARAMETERS 87

C1 Reading IRT Parameters 89
C2 Mathematics IRT Parameters 90
C3 Science IRT Parameters 91
C4 History/Citizenship/Geography IRT Parameters 92

APPENDIX D - TEST INFORMATION FUNCTIONS 93

D1 Reading 97
D2 Mathematics 98
D3 Science 99
D4 History/Citizenship/Geography 100

APPENDIX E - DESCRIPTION OF INDIVIDUAL ITEMS 101

E1 Reading 103
E2 Mathematics 105
E3 Science 107
E4 History/Citizenship/Geography 108

APPENDIX F - INTERCORRELATIONS OF TESTLETS 109

APPENDIX G - DEFINITIONS OF PROFICIENCY SCORES 113

APPENDIX H - STANDARD ERRORS OF MEASUREMENT AT THETA SCALE POINTS 121

CHAPTER 1. INTRODUCTION

The National Education Longitudinal Study of 1988 (NELS:88) is designed to monitor the transition of a national sample of young adults as they progress from junior to senior high school and then on to postsecondary education and/or the world of work. The NELS:88 surveys are monitored by the Longitudinal and Household Studies Branch (LHSB) of the National Center for Education Statistics (NCES). NELS:88 is the third and most recent in a series of longitudinal studies that are designed to provide timely information on trends in academic achievement. The two earlier longitudinal studies sponsored by NCES were the National Longitudinal Study of the high school class of 1972 (NLS) and the High School and Beyond (HS&B) study of 1980.

The primary purpose of this longitudinal data collection effort is to provide policy-relevant information concerning the effectiveness of schools, curriculum paths, special programs, variations in curriculum content and/or mode of delivery in bringing about educational growth. Although similar in its purposes to its two predecessors (NLS-72 and HS&B), NELS:88 is more comprehensive in the amount and type of data collected, as well as in the time period spanned by the data collection.

The base year sample was composed of approximately 24,600 eighth grade students who were sampled from slightly more than 1000 schools in the spring of 1988. These students are being followed up in the tenth grade (first follow-up) in the spring of 1990. The second follow-up will take place in the spring of 1992, which would normally be their senior year in high school. Attempts will be made to locate and survey sample members who have left school by that time or are not high school seniors. Post-secondary follow-up surveys are also being planned.

Among the more important educational indicators that will be monitored by the NELS:88 surveys is the achievement test battery. The NELS:88 test battery is critical for the measurement of academic growth that takes place between the eighth, tenth, and twelfth grades. In addition to measuring longitudinal growth during these critical years the NELS:88 battery will also be used to compare the performance of the NELS:88 sophomores in 1990 with the comparable 1980 sophomore cohort from the HS&B data collection, and 1992 NELS:88 seniors with the performance of HS&B and NLS-72 seniors.

For sample and race/ethnicity definitions and for detailed information about response rates, weighting, sample exclusions and survey methodology, please see the Base Year Student User's Manual (Ingels et al, 1990) and the Base Year Sample Design Report (Spencer et al, 1990).

The purpose of this report is to provide an in-depth description of the rationale, development, and subsequent statistical analysis of the eighth grade NELS:88 test battery.

CHAPTER 2. TEST SPECIFICATIONS

Aims and Objectives

The test specifications of the NELS:88 longitudinal test battery are dictated by its primary purpose--accurate measurement of the status of **individuals** at a given point in time as well as their growth over time. Like its predecessor, the 1980 High School and Beyond (HS&B) test battery, the National Educational Longitudinal Study (NELS:88) test battery was developed to measure both individual status and growth in a number of achievement areas. The four achievement areas are Mathematics, Reading Comprehension, Science, and History/Citizenship/Geography. However, unlike the HS&B assessment which was designed only to measure growth between the tenth and twelfth grades, the NELS:88 battery is designed to measure growth in achievement between the eighth, tenth and twelfth grades. Since the NELS:88 assessment spans four years with repeated testing of the same student cohort in the eighth, tenth and twelfth grades, it calls for a more flexible testing approach than was required in the HS&B longitudinal assessment.

The construction of the NELS eighth grade battery is in some sense a delicate balancing act between several competing objectives. Many of these objectives were suggested by the NELS Technical Review Panel (TRP) and/or NCES project staff during the base year development. Some of these objectives were as follows:

1) That the NELS:88 test battery cover four content areas - Reading, Mathematics, Science, and History/Citizenship/Geography.

2) That there be sufficient common items in the tenth grade mathematics form to link with the tenth grade 1980 HS&B cohort. Since the NELS:88 eighth grade mathematics test must also be linked to the tenth grade followup test, it would seem reasonable to have the linking items from HS&B be common to both the eighth and tenth grade NELS:88 mathematics tests.

3) That there be sufficient item overlap between the National Assessment of Educational Progress (NAEP) mathematics test and the eighth grade NELS:88 mathematics test to cross-walk to the NAEP mathematics scale if desired. Similar overlap was suggested for the NELS:88 reading test.

4) That the reading test passages provide relatively broad content coverage and have items that span at least three cognitive process areas. There also should be at least one passage that identifies in some way with minority concerns. Similarly, there should be at least one passage in which the main character is a female.

5) The Technical Review Panel suggested that the mathematics test, where possible, should emphasize concept understanding and problem solving skills in the areas of arithmetic, algebra, and geometry. It was felt that in a building block discipline such as mathematics, knowledge of the concepts that form the foundations that are later built upon are less likely to be learned and then forgotten.

6) The four content areas Reading, Mathematics, Science, and History/Citizenship/ Geography must be administered (including time for administration instructions) within one hour and a half.

7) The tests should be sufficiently reliable to support change measurement, and in the case of mathematics and reading be characterized by a sufficiently dominant underlying factor to support the Item Response Theory (IRT) model. This latter requirement is necessary to support the vertical equating between retestings as well as the cross-sectional linking with HS&B and NAEP, if desired. Given the time constraints, this is a "tall order". In order to achieve this level of reliability, as well as reduce the possibility of "floor and ceiling" effects, the Mathematics and Reading tests will be designed to be multi-level at the tenth grade.

Two-Stage Testing in a Longitudinal Framework

The potentially large variation in student growth trajectories over a four year period argues for a longitudinal "tailored testing" approach to assessment. That is, in order to accurately assess a student's status both at a given point in time as well as over time, the individual tests must be capable of measuring across a broad range of ability/achievement. If the **same** test, in say, Mathematics and Reading Comprehension were administered to the same student at the eighth, tenth, and twelfth grades the potential for observing "floor effects" at grade eight and "ceiling effects" at grade twelve is greatly increased. Of course if all four tests were quite long and included many very difficult as well as many very easy items, then theoretically there would be little opportunity for floor and ceiling effects to operate.

Unfortunately operational versions of the test must be relatively short in order to minimize the testing time burden on the students and their school systems. One potential solution to this problem is to use a two-stage testing procedure that allows one to at least partially tailor a test form to a particular individual's ability/achievement level.

That is, a two-stage longitudinal testing procedure will be implemented that would use the eighth grade test results for each student to assign him or her to a different form of the test when he or she is re-tested in tenth grade. For example, students scoring relatively high on the eighth grade test, in say, mathematics would be given a more difficult mathematics test form when they are retested as tenth graders. Students scoring relatively low in the eighth grade would receive an easier form when retested as tenth graders. Since tenth grade students would be taking forms that were in a sense appropriate to their particular level of ability/achievement, measurement accuracy would be enhanced and floor and ceiling effects would be minimized. The relative absence of ceiling effects should make the assessment of gain more accurate for students who had relatively high scores as eighth graders. Similarly, an accurate estimate of gain for low scoring eighth graders should also be enhanced, since floor effects should be minimized.

What does the utilization of a two-stage procedure have to say about how the components of the NELS:88 **eighth grade** battery should be constructed? Since at least some of the eighth grade tests (reading and mathematics) are to serve as "branching" or "routing" tests, ideally they should have good measurement properties throughout the test score range. That is, the test scores should provide reliable information at both the high and the low end of the test score distribution since students in these score ranges will be routed to tests of quite different average difficulties in the tenth grade.

Difficulty Level

The eighth grade reading, mathematics, and to a lesser extent the science and history/citizenship/geography tests were designed with these broad band measurement properties in mind. Operationally the goal of maintaining good measurement accuracy throughout the test score range is accomplished by building tests with a relatively rectangular frequency distribution of item difficulties. The typical test tends to follow a normal distribution of difficulties with the majority of the items in the middle difficulty range. A normal distribution of difficulties is considered to be relatively optimal if:

1) The population being tested is relatively homogeneous with respect to the ability/achievement being measured.

2) Diagnostic decisions (e.g., routing to different second stage tests) need not be made for individuals at either the high or low end of the test score (ability) distributions.

3) Reliable measurement of status at a given point in time is of primary importance and not the measurement of change. Ideally, change score analysis should be able to model a developmental growth model that has students at different points along the growth trajectory. If a test is built to simulate the various points along the growth trajectory, i.e., some items are selected for inclusion based on how well they represent steps in the developmental growth model, then there needs to be a greater diversity of item difficulties. Items should not all be "packed" at the middle difficulty level since that at best could only reflect accurate measurement of one step in the underlying developmental model.

4) Students are grouped into homogeneous ability/achievement groups based on say, a previously administered routing test. Students then could be administered separate test forms with each form having the majority of its items at the appropriate difficulty level for the corresponding ability grouping.

At the eighth grade level the total population is relatively heterogeneous. In addition, as pointed out above, the present plans call for the tenth grade students to be routed to different test forms depending on how well they did on their eighth grade testing. Separate mathematics and reading forms varying in average difficulty will be administered to homogeneous groupings of students based on their eighth grade achievement scores. These "tailored" test forms will be more homogeneous with respect to item difficulties within a test form since they are designed to match the ability level

of the test taker. However, since one of the purposes of the NELS:88 eighth grade battery is to provide diagnostic or routing information for the succeeding administration in the tenth grade, we have emphasized a broader range of item difficulties in the eighth grade tests.

IRT Scaling for Longitudinal Measurement and Equating to Earlier Cohorts

In order to accurately measure the extent of eighth to tenth grade gains at both the group and individual level, the eighth grade tests and the various forms of the tenth grade tests must be calibrated on the same scale. The most convenient way of doing this is to use Item Response Theory (IRT). In order to successfully carry out such a calibration for, say mathematics and reading, both the eighth and tenth grade tests should be relatively unifactorial with the same factor underlying both test administrations. This suggests that there be a common set of anchor items across eighth and tenth grade forms, and that most, but not necessarily all, content areas be represented in both eighth and tenth grade forms. Increments in difficulty demanded by future tenth and twelfth grade forms can be accomplished by: (1) increasing the problem-solving demands within the same familiar content areas and (2) including content in the later forms that tap materials normally found in the advanced course sequence.

The NELS:88 test battery scores must not only be put on the same vertical scales (i.e. from eighth to tenth to twelfth grade) but the mathematics items administered in the tenth grade must also provide "anchors" to the tenth grade HS&B mathematics items administered in 1980. While not required by contract, it would be desirable to be able to cross-walk the 1980 HS&B sophomore reading scores to the 1990 NELS:88 sophomore reading scores. The ability to put both the HS&B and NELS:88 sophomores on the same scale allows for a 10 year span cross-sectional trend comparison as well as the potential for a 10 year comparison between the HS&B sophomore to senior gains in 1980-1982 vs. those made by the NELS:88 students between 1990 and 1992. Appropriate use of IRT-scaling for these purposes requires that, to the extent possible, the tests be single-factor.

This cross-sectional scaling in addition to the vertical scaling (eighth through twelfth) puts additional constraints on mathematics and reading item selection for both the eighth grade and the subsequent follow-up tests. That is, in the case of mathematics at least 10 to 12 of the items should be common to both the eighth and tenth grade NELS:88 battery as well as to the tenth grade HS&B battery.

Psychometric Goals of the NELS:88 Eighth Grade Test Battery

While the long-term purpose of the NELS:88 battery is to accurately measure the status and growth of students at the individual level in four broad achievement areas, there are a number of allied psychometric and policy concerns that need to be addressed in the eighth grade battery. These concerns are as follows:

- Item selection should be curriculum-relevant, with emphasis on concepts, skills and general principles. When measuring change or developmental growth, the overemphasis on isolated facts at the expense of conceptual and/or problem-solving skills may lead to distortions in the gain scores due to forgetting. More will be said about this later.

- The tests should be relatively unspeeded with the vast majority of students completing all tests.

- There should be little evidence of floor or ceiling effects if the same test is to be repeated in the tenth grade.

- Reliabilities of the component tests should be psychometrically acceptable for the purpose of measuring **individual** status as well as growth. Unlike NAEP, which only assesses the status of groups, the NELS:88 battery must assess individuals and thus the tests require proportionately greater reliability than do their NAEP counterparts.

- The accuracy of measurement, i.e., the standard error of measurement, should be relatively constant across SES, sex and racial/ethnic groups. In fact, the NELS:88 battery is specifically designed to reduce the gap in reliabilities that is typically found between the majority group and the racial/ethnic minority groups.

- The test components should demonstrate some discriminant validity. That is, while the tests should be internally consistent and essentially be unifactorial (in the case of Reading and Mathematics), they should yield a relatively "clean" although oblique four factor solution. The four factors should be defined by the four tested content areas.

- Subscores and/or proficiency scores should be provided where psychometrically justified. The test specifications were designed to provide behaviorally-anchored proficiency scores in the areas of Mathematics and Reading.

- The NELS:88 test battery should attempt to minimize Differential Item Functioning (DIF) across gender and racial/ethnic groups that arises from irrelevant content that favors one or more of the groups. This, of course, refers to the so-called item bias problem.

- The NELS:88 test battery should share sufficient common items both across grade levels and with the HS&B battery to provide articulation of scores for vertical equating in NELS:88 as well as cross-sectional equating with HS&B.

Many of the following analysis results address the above concerns.

Specifications for Individual Tests

Given that the maximum allowable testing time for eighth graders was approximately one hour and thirty minutes, it was decided that the time would be apportioned in the following way among the test battery components:

Reading - Twenty-one questions in twenty-one minutes.
Mathematics - Forty questions in thirty minutes.
Science - Twenty-five questions in twenty minutes.
History/Citizenship/Geography - Thirty questions in fourteen minutes.

Based on simulations utilizing field test results (Rock & Pollack, 1987), ETS test development experts felt that these separately timed content areas would provide accurate assessment of each content area while minimizing any speededness component. The items that were used in the final eighth grade forms were selected from a much larger pool of items composed of items from NAEP, HS&B, the Second International Mathematics Study (SIMS), ETS test files from previous operational tests, and a pool of items specifically written for the NELS:88 Battery. The selection of items for the pre-test item pools was based on the consensus of the members of subject matter committees made up of curriculum experts. The subject matter committees consisted of educators, teachers, and college professors specializing in middle school curricula. There was considerable personnel overlap with similar subject matter committees used in the NAEP item pool development. ETS test development specialists were in attendance and worked with their respective subject matter committees in developing the eighth grade assessment objectives. Once the assessment objectives were agreed upon the subject matter committee members classified the items according to the objectives. A pool of 50 Reading items, 82 Mathematics items, 42 Science items, and 60 History/Citizenship/Geography items was selected for pretesting. Field tests were administered to eighth, tenth and twelfth graders in the Spring of 1987 (Rock & Pollack, 1987). The results of the field testing were scrutinized by additional committees of subject matter experts who suggested numerous modifications in content, format and wording of the items, as well as making judgments on content coverage. Final revisions and item selections were made by project staff on the basis of their input, and reviewed by NCES staff.

The following sections contain descriptions of the content and format of each of the four achievement tests. More detailed item-by-item specifications of the curriculum content, cognitive process, format, source, and particular content of the test items can be found in Appendix E.

Reading

The reading test consisted of five reading passages, ranging in length from a single paragraph to a half-page. Each passage was followed by three to five multiple choice

questions addressing the students' ability to reproduce details of the text, translate verbal statements into concepts (comprehension), or draw conclusions based on the material presented (inference/evaluation). A total of 21 questions were presented in 21 minutes. The amount of time allowed for each question, which is relatively long compared to the other three content areas, takes into account the length of time needed for reading the passages before answering the questions.

The reading test began with the least difficult (literary) passage followed by five relatively easy questions. The percent answering each item correctly (P+ a measure of item difficulty) by total and subgroups is presented in Appendix A-1. The next passage was a short science passage followed by three questions. These three questions were more difficult than those associated with the literary passage. The increased difficulty could be due to the science content or the fact that the questions went beyond simple reproduction of detail. The next passage was a six item poetry passage. The item difficulties varied from relatively easy to relatively difficult. The fourth passage was a biographical piece concerning the Black jazz musician Louis Armstrong and was followed by four questions of medium difficulty. The last three items were based on a passage discussing the role of pioneer women. These items were relatively easy. The first eight items in the reading test used a five option multiple choice format while the remaining fifteen items used a four option multiple choice format. Other than to present a relatively easy passage first no conscious attempt was made to present the remaining items in order of difficulty. The motivation for including several very easy items on this test came from the field test results. Pretesting of the reading materials indicated the possibility for floor effects for some individuals.

Figure 1 presents a two-way table of reading passage content categories by cognitive process categories for the reading test. The entries in the cells of the matrix are the number of items in that particular cross-classification. Appendix E-1 contains additional details on the content and characteristics of individual items.

Inspection of Figure 1 indicates that the eighth grade test attempted to cover as many content areas as possible given the limitations inherent in the time allocation. In order to achieve a reasonable level of discrimination for the low, middle and higher level readers, there were items requiring simple reproduction of detail as well as items requiring comprehension and inference skills. One passage (the biographical passage) discussed the life of a Black musician. The primary characters in one of the other passages were women pioneers. The remaining passages did not contain references to the race/ethnicity of the characters, and the gender of the characters was not an important issue. This attempt to balance the content of the reading passage with respect to gender and race/ethnicity represents an effort to reduce the potential for bias affecting subgroups of the population.

As expected, the comprehension and inference/evaluation items tended to be somewhat more difficult than those items requiring simple reproduction of detail. While the comprehension and inference/evaluation items were more difficult on average than the reproduction of detail items, they were purposely designed not to be extremely difficult for the typical eighth grader for two reasons:

Figure 1.--Reading test specifications (number of items by process and content)

	CONTENT			
PROCESS	Literary	Science	Poetry	Biography
Reproduction of detail	3	1	-	-
Comprehension	-	1	1	1
Inference and/or Evaluation	5	1	5	3

1) We were not concerned about ceiling effects at grade 8 imposing artificial constraints on eighth to tenth grade gains since we were planning to route students to forms that would be appropriate for their ability level at the tenth grade.

2) We were attempting to increase the accuracy of measurement for the low SES and/or racial/ethnic groups who traditionally score lower on cognitive measures. The trick is to accomplish this goal without sacrificing the overall reliability, i.e., the reliability estimated for the total population. Widening the range of item difficulties to include several very easy items was intended to aid in reaching this objective.

Mathematics

The proportion correct (P+) for the mathematics test items are presented in Appendix A-2. The first 19 items in the mathematics test are referred to as quantitative comparison items. While these items follow the multiple choice mode they have a somewhat different format than the typical multiple choice item. The student is presented with two quantities--one in column A and one in column B. He or she is then asked to compare the two quantities and mark option (A) if the quantity in column A is greater; (B) if the quantity in column B is greater; (C) if the two quantities are equal; and (D) if the relationship cannot be determined from the information given.

These first 19 quantitative comparison items cut across most of the content areas but tended to be classified as skills and/or declarative knowledge or understanding/ comprehension of concept. The quantitative comparison item type was included in the mathematics test for two reasons. First and primarily, this was the only item type used in the HS&B mathematics test and thus they can provide us with the common item anchors needed for the cross-sectional equating. Secondly they tend to take less time to administer than other formats and thus the student can do approximately three quantitative comparison items for every two standard multiple choice items. Assuming equal item reliabilities we can achieve significantly higher test reliability for a fixed amount of testing time. Inspection of the item biserials (a measure of an item's reliability) in Appendix A-2 does suggest that the item reliabilities of the quantitative comparison and the standard multiple choice are about the same.

One additional concern about the quantitative comparison item types is that the format might be sufficiently unfamiliar to some of the students to make them artificially difficult. Inspection of the item difficulties in Appendix A-2 suggest that they appear to run the gamut from easy to hard. The finding that they are not differentially difficult for minority groups will be treated in the section dealing with differential item performance.

The remaining mathematics items are the standard 4 option and 5 option multiple choice items types, containing a mix of word problems, diagrams, and calculations. There is a slight ordering with respect to difficulty since the more difficult problem solving items were placed near the end of the test.

Figure 2 presents the test specifications in terms of item classifications for the eighth grade mathematics test. See Appendix E-2 for content information on an item-by-item basis.

Inspection of Figure 2 indicates that nearly half of the of items in the eighth grade mathematics test can be classified as requiring skills or declarative knowledge. The "skills and declarative knowledge" category actually includes two relatively separable

Figure 2.--Mathematics specifications (number of items by process and content)

	CONTENT				
PROCESS	Arithmetic	Algebra	Geometry	Data/ Probability	Advanced Topics
Skills/ Knowledge	10	4	1	1	1
Understanding/ Comprehension	6	7	3	3	-
Problem Solving	3	-	-	-	1

knowledge demand levels. The lowest level consists primarily of simple arithmetical operations on whole numbers and the second level requires skills in operations with decimals, fractions, and percentages. The "understanding/comprehension" level consists of items that require translating verbal statements and concepts into figures, and demonstrating understanding of concepts and principles through explanation, recognition or illustration. For example, arrival at the correct answer may involve **understanding** the relationship between decimals and percentages, etc. The higher order problem solving category is less well defined at this level (eighth grade) but it typically involves generalizing and applying mathematical knowledge, skill and comprehension in situations requiring reasoning, judgment, and decision-making processes. It is anticipated that the tenth grade mathematics forms will include a larger representation of items requiring problem solving skills.

It should be pointed out here that when one computes content subscores based on say, the arithmetic and algebra items, one should not be surprised if such subscores are very highly correlated since both content areas include similar item distributions with respect to cognitive demands (i.e., processing demands). Most students, by the eighth grade, have been exposed to instruction in the skills needed to solve the **lowest level** (Skills/Knowledge) items. Therefore, individual differences in performance are going to be driven by differential exposure and practice in the **higher-level** skills related to concept understanding and simple problem solving.

Subscores or proficiency scores based on the rows (cognitive processes) of the above classification matrix may have a greater potential for discriminable subscores than are the columns (Content areas). The rows that define the cognitive processes tend to follow a difficulty hierarchy. That is, the skills at each higher level require all the skills of the lower levels plus some new additional skill. This hierarchy in complexity tends to make subscores based on items describing these different cognitive process levels somewhat more differentiable than those based on the content areas. The increase in conceptual complexity as one goes from the simple rule-following of the declarative knowledge items to the item types representing conceptual understanding and finally problem solving, suggest that possibly qualitatively different skills come into play as one proceeds up the "ladder" of complexity.

Science

The item format for the science test is the standard multiple choice format with approximately two-thirds being four choice and the remaining items five choice. The majority of the items contain a verbal description of a situation followed by a question based on the premise. Several items include graphs or diagrams illustrating the circumstances described. There is a considerably stronger relationship between item sequence and item difficulty in the science test when compared to the reading and mathematics tests. That is, inspection of Appendix A-3 indicates that there is a relatively consistent increase in item difficulty as one proceeds from the beginning to the end of the test. Indeed the science items were ordered to reflect their pretest difficulties.

Figure 3 presents a two-way table of the classification of the Science items. Additional detail on characteristics and content of individual items can be found in Appendix E-3.

Since no computations are involved in the science items (unlike the higher level mathematics items) and inferences from facts may be more straightforward than in the reading comprehension test, often understanding the concept is tantamount to solving the item. As a result these process classifications in science are particularly sensitive to differences in opinion among science experts. Content areas in science also have a tendency to overlap with each other. While this is true for the other areas also, it is especially true for science items.

History/Citizenship/Geography

The History/Citizenship/Geography test items were only classified according to content area. Of the 30 items in the test, fourteen were history questions; thirteen were citizenship/government questions, and the remaining three items dealt with geography/ economic development.

Figure 3.--Science test specifications (number of items by process and content)

	CONTENT			
PROCESS	Earth	Life	Chemistry	Scientific Method
Declarative Knowledge	5	3	2	
Comprehension	2	2	2	1
Problem Solving	1	3	3	1

The three content areas were distributed throughout the test. The items were sequenced for the most part on the basis of their pre-test difficulties with the easier items in the beginning and the most difficult items near the end. Appendix A-4 presents the item difficulties. Content, source, and descriptive information on each item can be found in Appendix E-4. The item format consisted of twenty-two four option multiple choice with three five option multiple choice and five true-false items.

Matching Test Content to Curriculum

The question of overlap between test items and curriculum content has received increasing attention over the last ten years and evaluation methodologies have come to be dominated by the **doctrine of maximal overlap** (Frechtling, 1989). Mehrens (1984) and Cronbach (1963), however, questioned whether maximal overlap is in fact desirable except possibly in those cases where a specific program is being evaluated. Mehrens argues that a close match between curricular and test content is desirable only if one wishes to make inferences about specific objectives taught by a specific teacher to a specific school. Even if one would wish to evaluate the effects of a specific teacher in a specific class, one inference of importance is the degree to which the specific knowledge taught in that class generalizes to other relevant domains.

Nitko (1989) argues that tests designed to measure individuals and to facilitate their learning within a particular instructional context are not necessarily optimum for measuring school or program differences. Similarly Airasian & Madaus (1983) suggest that the following design variables be taken into account:

(A) The ability of tests to detect differences between groups of students.

(B) The relative representativeness of the content-behavior-process sampled by test items.

(C) The parallelism of the response formats and mental processes learned during instruction with those defined by the test tasks.

(D) The properties of the scores and the way that they will be summarized and reported.

(E) The validity of the inferences about school and program effectiveness that can be made from the test results.

Experience and practice suggests that tests are **unlikely** to detect differences between schools and programs when total test scores are used and when the subject matter tested is likely to be related to learning in the home (e.g. reading) rather than to schooling (e.g. mathematics) (Airasian & Madaus, 1983; Linn & Harnisch, 1981).

Schmidt (1983) identifies three major types of domains from which content to be covered can be drawn: a priori domains, curriculum-specific or learning-material-specific domains, and instructional material domains. Nitko (1983) suggests that "agents" not

associated with local schools or particular programs tend to define a priori domains by using social criteria in judging what is important for all to learn. He goes on to suggest that test exercises in the National Assessment of Educational Progress (NAEP) as well as state assessment programs are examples of assessment instruments built from a priori domains since they specify content to be included without linking that content to specific instructional material or specific instructional events.

Cole & Nitko (1981) suggest that another design variable be considered in building tests to detect school and program effectiveness. They suggest that students require more time to acquire global skills and to grow in general educational development than to learn specific knowledges and skills. They suggest that tests measuring the former are less sensitive to measuring short term instructional efforts than tests measuring the latter.

Cooley (1977) and Leinhardt (1980) argue for the collection of relevant classroom variables and developing tests that are sensitive to differences between classrooms within-program. Leinhardt & Seewald (1981) describe several within-school, program, and classroom variables that are important to program evaluators and how to measure them. Mehrens and Phillips (Mehrens, 1984; Mehrens & Phillips, 1986; Phillips & Mehrens, 1988), however, found no significant differences on standardized tests from the use of different textbooks and different degrees of curriculum-test overlap when previous achievement and socioeconomic status were taken into account.

What we have attempted to do here is take kind of a middle road in the sense that our curriculum experts were instructed to select items that were curriculum relevant but typically did not require a great deal of isolated factual knowledge. The emphasis was to be on understanding concepts and the measurement of problem-solving skills. However, it was thought necessary to assess the basic operational skills (e.g., simple arithmetic and algebraic operations) which are the foundations for successfully carrying out the problem solving tasks.

The incorporation in the mathematics test of the relatively simple arithmetic and algebraic items which measure procedural or factual knowledges served two purposes. First, this subset of items provided better assessment for those low scoring students who were just beginning to develop their "basic mathematical skills". Second, these items should be able to provide a limited amount of diagnostic information about why some students are not able to successfully carry out the tasks defined in the typically more demanding problem solving items. For example, students who are not proficient on the problem solving items can be further divided into two groups based on their performance on the arithmetical/algebraic procedural skill items. One subgroup could not very well be proficient on the problem solving items since they did not demonstrate sufficient skills on the simple arithmetical/algebraic procedures that are a necessary but not a sufficient condition for successful performance on the problem solving tasks. The remaining subgroup, however, had sufficient grounding in the basics as demonstrated by their successful performance on the procedural items but were unable to carry out the logical operations necessary to complete the solutions to the problem solving items.

This hierarchical nature of the required skills is put to formal use in the development of behaviorally anchored proficiency level scales for both reading and mathematics. This criterion referenced interpretation is discussed further on under the subtopic Proficiency Level Subscores.

This concern with respect to the maximal overlap doctrine is particularly relevant to the measurement of change over relatively long periods of exposure to varied educational treatments. That is, the two year gaps between re-testings coupled with a very heterogeneous student population are quite likely to coincide with considerable variability in course taking experiences. This fact, along with the constraints on testing time, makes coverage of specific curriculum related knowledges very difficult. Also, as indicated above, specificity in the knowledges being tapped by the cognitive tests could lead to distortions in the gain scores due to forgetting of specific details. It is our opinion that the impact on gain scores due to forgetting will be minimized if the cognitive battery increasingly emphasizes general concepts and development of problem solving abilities. This emphasis should increase as one goes to the tenth and twelfth grades. Students who take more high level courses, regardless of the specific course content, are likely to increase their conceptual understanding as well as gain additional practice in problem solving skills.

At best any nationally based longitudinal achievement testing program must be a compromise that best attempts to balance testing time burdens, the natural tensions between local curriculum emphasis and more general mastery objectives, and the psychometric constraints (in the NELS:88 case) in carrying out both vertical equating and cross-sectional equating. NELS:88 fortunately does have the luxury of being able to gather longitudinal pre-test data on the item pools. Thus we have been able to take into consideration not only the curriculum relevance but whether or not the items demonstrate reasonable growth curves, as well as meet the usual item analysis parameter requirements for item quality.

CHAPTER 3. PSYCHOMETRIC ANALYSIS RESULTS

Were the Tests Speeded?

ETS uses a two-part "rule-of-thumb" for determining whether or not a test is speeded. A test is considered to be unspeeded if nearly all test-takers reached the three-quarters point of the test, and at least 80 percent of the students answered the last item. The first criterion was met by 97 percent or more of students in all subgroups for all four NELS:88 tests, with the exception of Black students, 95 percent of whom reached the three-quarters point on the reading test. Table 1 below presents the statistics for the second criterion, percent answering the last item. Inspection of the entries in Table 1 indicate that all tests exceeded this criterion by a considerable margin for all groups. In a test such as NELS:88, which represents a "no risk" situation for the student, failure to answer items may be due to a lack of motivation as well as to insufficient time. It is evident that the allocated test timings were appropriate for all eighth grade groups.

Table 1.--Speededness indices for tests, by racial/ethnic and sex groups (percent of sample who reached last item)

TEST	Asian	Hispanic	Black	White	Male	Female
Reading	96.1	92.7	87.9	97.3	94.9	95.9
Math	96.1	93.2	89.7	96.2	95.0	94.9
Science	96.2	95.3	92.6	98.0	96.7	97.0
Hist./Citiz.	96.6	95.5	94.6	97.9	97.0	97.3

SOURCE: U.S. Department of Education, National Center for Education Statistics, NELS:88 Base Year Survey.

Reliabilities of the NELS:88 Eighth Grade Test Battery

Table 2 presents the reliabilities and standard errors of measurement for racial/ethnic and sex groups for each test in the NELS:88 eighth grade battery. These reliabilities are based on weighted data. For comparison purposes the reliabilities and

standard errors of measurement are also shown for the analogous components of the HS&B sophomore test battery (Rock et al., 1985). The reliabilities are internal consistency measures based on coefficient Alpha. High coefficient Alpha reliabilities (eighties and above for tests of this length) suggest that the tests are relatively unifactorial. While standard errors of measurement (SEM's) are presented for both the NELS:88 and the HS&B battery, they (the SEM's) are not strictly comparable, since both the instruments **and** the populations are different. In such cases, reliabilities are the preferred measure of accuracy.

The results in Table 2 suggest that the reading and math tests in the NELS:88 battery provided an increment in reliability over that provided by their counterparts in the HS&B battery. This increment in reliability is particularly noticeable in the reading area and to a somewhat lesser extent in mathematics. The large gains in reliability in these two content areas are particularly welcome since they seem to be greatest for the minority populations. It was hoped that the reliabilities of the traditionally lower scoring groups, e.g., Blacks and Hispanics, could be increased without an accompanying decrease for the white majority. As indicated earlier one of the test construction goals in mathematics and reading was to provide a more rectangular distribution of difficulties across the low and middle difficulty levels, thereby providing additional discrimination at the low end of the test score distribution.

One should keep in mind here that we are comparing different populations. A more accurate summary of Table 2 is that the NELS:88 reading and mathematics tests do a better job of assessing eighth graders than did the comparable tests in the HS&B battery when administered to tenth graders. It should also be pointed out that the NELS:88 mathematics test included two more items than did its counterpart in HS&B. Similarly, the NELS:88 reading test had one more item than did its counterpart in HS&B. These differences in numbers of items are not of sufficient size to completely explain the gains in reliability. The increased overall reliability (i.e., for the total sample) is more likely to have resulted from the fact that the test specifications took into consideration the intention of tailoring the tenth grade follow-up test forms (at least in reading and mathematics) to the ability of the students as described by their eighth grade scores. That is, since the eighth grade test was not intended to be re-used at tenth grade, it could be constructed to best measure the range of achievement expected in the base year without concern for potential ceiling effects later on. HS&B used the same test forms to measure students in both tenth and twelfth grades. This implies some compromises in test specifications, a constraint which was not in effect in designing the NELS:88 tests.

Knowing that we were intending to change the tenth grade test allowed the test developers to build an eighth grade test that only needed to maximize the accuracy of assessment at the eighth grade. If the test development project staff had been directed to build a reading and mathematics form that was to be the same for both eighth and tenth graders, then the final eighth grade form would have been more difficult on average in order to minimize ceiling effects at the tenth grade level. The increased difficulty would, of course, tend to reduce the reliability of the eighth grade test, particularly for the low scoring individuals.

Table 2.--Test reliabilities and standard errors of measurement (in parentheses), by race/ethnicity and sex

	Asian	Hispanic	Black	White	Male	Female	TOTAL
				READING			
NELS:88 Rel	.85	.79	.77	.83	.84	.83	.84
NELS:88 SEM	(2.43)	(2.57)	(2.60)	(2.47)	(2.48)	(2.48)	(2.46)
HS&B Rel	-	.64	.66	.76	.77	.76	.77
HSB SEM	-	(2.30)	(2.23)	(2.28)	(2.29)	(2.27)	(2.28)
				MATHEMATICS			
NELS:88 REL	.92	.86	.84	.89	.90	.90	.90
NELS:88 SEM	(3.46)	(3.70)	(3.62)	(3.66)	(3.62)	(3.53)	(3.57)
HSB REL	-	.79	.76	.87	.88	.85	.87
HSB SEM	-	(3.57)	(3.51)	(3.51)	(3.51)	(3.53)	(3.52)
				SCIENCE			
NELS:88 REL	.77	.67	.62	.74	.78	.72	.75
NELS:88 SEM	(2.89)	(2.98)	(2.96)	(2.90)	(2.86)	(2.92)	(2.91)
HSB REL	-	.68	.64	.69	.76	.71	.74
HSB SEM	-	(2.44)	(2.40)	(2.33)	(2.32)	(2.40)	(2.36)
				History/Citizenship/Geography			
NELS:88 REL	.86	.81	.76	.83	.85	.82	.83
NELS:88 SEM	(3.03)	(3.33)	(3.38)	(3.01)	(3.06)	(3.10)	(3.15)

- No Comparable test in the HS&B Battery-

SOURCE: U.S. Department of Education, National Center for Education Statistics, NELS:88 Base Year Survey and High School and Beyond Base Year Survey.

It was encouraging to observe that the eighth grade NELS:88 Science test achieved about the same degree of reliability as the tenth grade HS&B test. One would not expect many eighth graders to be exposed at this point in their development to some of the material in the Science test. Given the number of life and earth science items and to a lesser extent chemistry items, it is believed that the test will be more appropriate when given to tenth graders who will have been exposed to additional coursework in these areas, and thus should show additional incremental gains in measurement accuracy at that point in time.

Similar to the Reading and Mathematics test, the History/Citizenship/Geography (HCG) test also demonstrated relatively high internal consistency reliability. The internal consistency reliability of the HCG test was sufficiently high to suggest that IRT methods could be used to put more than one form on the same scale if required in the follow-ups. Inspection of histograms and p-plots for the HCG test suggest a slight ceiling effect if we used the same form again in the tenth grade.

A simple descriptive index of the potential for a ceiling effect is the difference between the mean and a perfect score divided by the standard deviation. If the distribution is relatively normal in the sample, then there should be slightly more than 2 standard deviations between the mean and a perfect score. In the case of the Science test this index is equal to 2.47, indicating almost two and a half standard deviations between the eighth grade mean and a perfect score. In addition, both histograms and p-plots of the Science scores suggest that the sample distribution more nearly approximates a normal distribution than that of any of the other tests.

The same index for the HCG test is equal to 1.87 suggesting that there is some potential for a ceiling effect here if the same form were used at the tenth grade. The results of the follow-up pretest (Rock & Pollack, 1989) also suggested the need for a vertically equated more difficult tenth grade form.

Originally both the Science and the HCG tests were considered to be candidates for keeping the same form at least through the tenth grade. There is little evidence arising from the eighth grade data that suggests that this may not be a viable way to go in the case of the Science test. Also using IRT methods for putting different forms of the Science test (e.g., different tenth & twelfth grade forms) on the same scale might be somewhat problematic because of the relatively low internal consistency of science items. Fortunately the HCG test appears to be sufficiently internally consistent for IRT scaling and thus there is the potential for including more difficult items in the tenth grade test.

Item Statistics by Gender and Racial/Ethnic Groups

Appendices A1-A4 present traditional item analysis statistics including the item difficulties (**P+**), item biserials, and deltas. The item difficulties are simply the proportion of students who passed a particular item. The item biserials are measures of the relationship between performance on a given item and on the total pool of items as measured by the total score. The item biserial is often considered to be a measure of

given item's reliability. Another way of looking at the biserial is that its size reflects the extent to which a given item measures the "same things" as the remainder of the test.

Items yielding biserials of .40 are considered to be quite reliable while items at .50 and above are considered to have excellent reliability. Items that have biserials in the 0-.20 range, or worse yet are negative, would be candidates for replacement.

The item deltas are defined as $\Delta = 4\ \Phi^{-1}(1-P_i^*) + 13$ where Φ^{-1} is the inverse normal transformation that transforms a probability value into a normal deviate with unit variance. Thus the distribution of item deltas will have a mean delta of 13 and a standard deviation of 4. Item deltas are used by ETS test development specialists as the index of item difficulty in defining test specifications.

In Appendices A1-A4, at the bottom of each column are summary statistics for the item analysis. The item biserials for the NELS:88 battery are all positive and relatively high for all groups. There is, however, a consistent tendency for the biserials to be somewhat lower for the Hispanics, Blacks, and American Indians. This is at least partly an artifact of the slightly lower total test score variances for these groups. Table 3 below summarizes the item difficulty and biserial information by content area and compares these with their counterparts from the HS&B tenth grade data. As expected, the average biserial was somewhat higher for the NELS:88 reading and mathematics tests than for their counterparts in the HS&B battery. This finding is consistent with the higher reliabilities reported above for the NELS:88 reading and mathematics tests.

The fact that on average the NELS:88 reading and mathematics tests were somewhat easier than their HS&B counterparts (i.e., higher average P+) was also consistent with the design specifications that attempted to increase the reliability for the traditionally lower scoring groups. That is, the NELS:88 reading and mathematics tests had proportionately more easy items than did the HS&B battery. The larger number of easy items minimized the possibility of observing "floor effects" for the low scoring groups. As indicated above, the eighth grade test specifications were less driven by concerns about ceiling effects in the later followups than was the case for HS&B, since different and more difficult forms would be introduced at the tenth grade for NELS.

Unlike the reading and mathematics content areas, the science area was slightly more difficult for eighth graders than the comparable test for the HS&B tenth graders. This was anticipated since many eighth grade students probably had little familiarity with some of the content in the Science test.

Compared to the remaining tests in the NELS:88 battery, the average difficulty of the HCG test items suggests that it was the easiest test. This result is, of course, consistent with the earlier finding of a potential ceiling effect if the same form were used again in the tenth grade.

Table 3.--A comparison of average difficulty and average biserials for comparable tests in the HS&B and NELS:88 test battery

	NELS:88 Eighth Grade Average		HS&B Tenth Grade Average	
	P+	Biserial	P+	Biserial
READING				
Asian	.63	.65	Not available	
Hispanic	.52	.57	.38	.48
Black	.49	.55	.37	.50
White	.65	.64	.52	.57
TOTAL	.61	.64	.48	.57
MATHEMATICS				
Asian	.61	.64	Not available	
Hispanic	.45	.51	.39	.44
Black	.41	.49	.36	.42
White	.58	.57	.53	.53
TOTAL	.54	.58	.49	.53
SCIENCE				
Asian	.56	.51	Not available	
Hispanic	.46	.43	.45	.48
Black	.42	.41	.41	.46
White	.57	.49	.59	.52
TOTAL	.53	.49	.55	.54
History/Citizenship/Geography				
Asian	.67	.62	No comparable test	
Hispanic	.56	.51		
Black	.54	.48		
White	.66	.59		
TOTAL	.63	.58		

SOURCE: U.S. Department of Education, National Center for Education Statistics, NELS:88 Base Year Survey and High School and Beyond Base Year Survey.

Differential Item Functioning (DIF)

Differential Item Functioning (DIF) as defined here attempts to identify those items showing an unexpectedly large difference in item performance between a focal group (e.g. Black students) and a reference group (e.g. White students) when the two groups are "blocked" or matched on their total score. It should be noted that any such strictly internal analysis, i.e., without an external criterion, cannot detect bias when that bias pervades all items in the test (Cole & Moss, 1989). It can only detect differences in the relationships among items that are anomalous in some group in relation to other items. In addition such approaches can only identify the items where there is unexpected differential **performance**, they cannot directly imply **bias**. A determination of bias implies not only that differential performance on the item is related to subgroup membership, but also that the difference is **unfairly** associated with subgroup membership. That is, the difference is due to an attribute **not related** to the construct being measured. As Cole & Moss (1989) point out, items so identified must still be interpreted in light of the intended meaning of the test scores before any conclusion of bias can be drawn.

The DIF program was developed at the Educational Testing Service (Holland and Thayer, 1986) and was based on the Mantel-Haenszel odds-ratio (Mantel and Haenszel, 1959) and its associated chi-square. Basically, the Mantel-Haenszel (M-H) procedure forms odds ratios from two-way frequency tables. In a twenty item test, 21 two-way tables and their associated odds-ratios can be formed for each item. There are potentially 21 of these tables for each item since there will be one table associated with each total score from 0-20. The first dimension of each table is groups, e.g., Whites vs. Blacks, and the remaining dimension is passing vs. failing on a given item. Thus the question that the M-H procedure addresses itself to is whether or not members of the reference group, e.g., Whites, who have the same total score as members of the focal group, e.g., Blacks, have the same likelihood of passing the item in question. While the M-H statistic looks at passing rates for two groups while controlling for total score, no assumption need be made about the shape of the total score distribution for either group.

The chi-square statistic associated with the M-H procedure tests whether the average odds-ratio across all 21 score levels differs from unity, i.e., equal likelihood of passing.

Three columns in the M-H tables are of particular interest. The first of these three columns is labeled "prob > Chi-sq" and it provides a statistical test of whether or not the average odds-ratio significantly departs from unity. If the probability in this column is .05 or less then one could say that there is statistical evidence for DIF on the item in question. The problem with this interpretation is two-fold. First, one is making a number of statistical tests, one for each item, and second, if there are two relatively large samples involved, statistical significance will be guaranteed.

Given these reservations the Educational Testing Service has developed an "effect size" estimate that is not sample size dependent. These effect sizes are in the column labeled MH D-DIF. Associated with the effect sizes is a letter code that ranges from "A" to "C". It is ETS's experience that effect sizes of 1.5 and above are practically significant. Effect sizes of this magnitude, and which are statistically significant, are labeled with a "C" . Test development experts can often inspect items that are characterized by such large DIF properties and in some cases be able to provide a reasonable explanation for the differential item functioning. This has not been the case for items in the A or B DIF categories. The negative sign on the M-H D-DIF column indicates that the DIF is favoring the reference group and is against the focal or target group (typically the minority group). The third and last column of interest is the column labeled impact. This column simply shows the raw differences in the P+'s when the focal group's P+ is subtracted from that of the reference group.

If DIF statistics have been obtained on pretested items, all "C" items will normally be replaced in construction of an operational test, unless they are needed to meet test specifications. This is done regardless of whether the group differences are related to the construct. Once a test has been administered, however, replacement of items is no longer an option; the only choice possible is whether to accept the questioned item or drop it from scoring. At this stage, it has been the policy of the Educational Testing Service to submit items having "C" level DIF statistics to a test development committee for review. If the committee can identify content that is likely to be unfamiliar to the subgroup in question **and** which is irrelevant to the skill being measured the item will typically be removed from the test score. However, if the identified source of difference is consistent with the construct being measured, or if no reason for the difference can be determined, the item is retained.

Appendices B1-B20 present the tables of differential item functioning which compares the base or reference group (Whites or males) with each of the racial/ethnic or female comparison groups. For each test content area there are five DIF tables. For example, Appendix B1 presents the contrast between Whites and Asians on each of the

reading items. Appendices B2-B4 present contrasts between Whites and Hispanics, Blacks, and American Indians respectively. B5 presents the contrast between male and female on the reading items. Appendices B6-B20 repeat the same contrasts for the remaining three content areas.

Inspection of the effect size columns suggest that there is little or no evidence for the presence of DIF in the NELS:88 test battery. In the case of reading there is only one "C" level item and its sign is positive indicating that the DIF is favoring the focal group (American Indians in this case). There are 116 items in the NELS:88 Battery and there are 580 DIF contrasts being made. Because of the large number of contrasts being tested we will emphasize those items that show DIF for two or more groups.

The only "C" level item in the reading test heavily favored American Indians over Whites. However, an artifact of the computational formulas in the DIF procedure is that easy items are much more likely to be identified as showing DIF than hard items.

Reading item 1, with a P+ of .96 for Whites and .95 for American Indians, was by far the easiest item in the whole test battery.

In the case of the mathematics test there were only two "C" level DIF items. Item 25 favored the Whites over the Black students and also favored the male students over the female students. Item 25 requires only simple arithmetical operations but the units are in centimeters. It is possible that both Black and female students may be somewhat less comfortable with the concept of centimeters as the units of measurement. Item 37 favored the reference group (Whites) when compared with the focal group (Asians). Item 37 is a low level problem solving geometry problem which uses the term "stick-lengths" in the stem. It is possible that this hyphenated word was confusing to some of the Asian students. Inspection of the item biserial for the Asian group (Appendix A2) indicates that it is quite high (.69) suggesting that it does appear to be quite reliable and is discriminating the high scoring Asians from the low scoring Asians.

As mentioned earlier in the discussion of the quantitative comparison items, there is some concern about the possibility that they might be unfair to minority groups on the basis of their potential lack of exposure to the item format. Inspection of the first nineteen items (the quantitative comparison items) in appendix B-6 indicates that there are no "C" level items among the quantitative comparison items for any focal group comparison. In terms of "B" level items, the Asians have two- one in favor of the focal and one in favor of the reference group. When the Hispanics are the focal group all the contrasts for the first nineteen items are at the "A" level (difference is small and/or not statistically significant) and most of those favor the focal group. There are two "B" level quantitative comparison items in the Black vs. White student comparison. In both cases the items favor the focal group (Black students) rather than the White reference group. The American Indian--White student comparison only showed "A" level contrasts. It would appear that there is no evidence for DIF among the quantitative comparison items.

The science test had only one "C" level item (item 14) and that appeared to favor White students over Black students. This item refers to the temperature of a mixture of two liquids. Subsequent review of this item by the test development committee came up with no insights on why this item showed DIF. As in previous examples of item DIF, this particular item had a respectable biserial (.50) for the Black students.

Item 21 seemed to favor male students over females. Question 21 deals with how the interaction of water temperature and that of the land generates a sea breeze at the beach. A review of the item failed to identify any gender linked problems.

The HCG test had 5 items that showed "C" levels of DIF. Of particular interest here was item 9 which showed DIF in favor of the White students when compared with the Asian students, Hispanic students, and the American Indian students. Item 9 asks the student whether "refusing to obey laws" is a way that American citizens can legally oppose laws or actions of officials. While the biserials are quite high for this item in all the subgroups in question, this item may be measuring an attitude towards protest rather

than knowledge of what is legal and what is not legal. This item is a reasonable candidate for replacement in the tenth grade test.

Item 14 also yielded "C" level DIF statistics in two reference - focal group comparisons. The interesting finding about this item is that it favored the focal groups (Asian and Hispanic students). Item 14 asks about regions of the world that "the greatest number of immigrants to the United States come from".

Three other HCG items were identified, but each affected only one subgroup and in each case the statistic passed the cutoff for "C" items by a relatively small amount. Reviewers did not identify how these items are unfairly related to subgroup membership.

Given the number of items and group contrasts one has to conclude that there was little differential item functioning in the eighth grade NELS:88 battery. This happy result is probably due to the extensive pre-review of the items by both the ETS project development staff as well as the NCES staff.

Factor Structure of the NELS:88 Eighth Grade Battery

The factor structure of the NELS:88 battery was examined from two different complementary perspectives. These two perspectives were:

- Convergent validity--This analysis addressed the question of whether or not items grouped by content into parcels would indeed define a common factor. For example, do four separately constructed mathematics item testlets consisting of arithmetic, algebra, geometry, and probability items respectively define a single mathematics factor? Similar content based item testlets were constructed as "factor markers" in each of the other three tested areas.

- Discriminant validity--This analysis complements the convergent validity question in that it examines whether or not the factors defined by their marker testlets have discriminant validity. That is, is a mathematics factor separable from a reading comprehension factor and also from a science factor, etc?

The use of testlets to mark or define factors rather than individual items is advantageous since they (testlets) yield relatively continuous scores and are inherently more reliable than single items.

This does not mean that other recently developed alternative methods using factor analysis of item responses (e.g. Bock, Gibbons, & Muraki, 1985) might not also be helpful here. While the Bock et al. Testfact program would in theory allow us to factor analyze at the item level, we have experienced considerable problems with convergence with item data sets of the size being analyzed here. An approximation to the Bock et al. factor solution at the item level is presented in a following section dealing with dimensionality at item response theory.

Five testlets, each one representing a different reading passage, were used to mark a potential reading comprehension factor. The five testlets were based on a literary passage, science passage, poetry passage, biographical passage, and a historical passage. Four testlets were assembled to mark a mathematics factor. The four mathematics testlets consisted of arithmetic, algebra, geometry, and probability items respectively. Similarly four marker testlets were assembled from the science items. These testlets were composed of earth science, life science, chemistry, and scientific method items respectively. Three HCG testlets were formed based on History, Citizenship/ Government, and Geography/Economic development items respectively.

The 16 testlets were analyzed using maximum likelihood procedures for the factor extraction stage. Four factors were then rotated to an oblique solution using the Promax procedure (Hendricksen & White, 1964). Table 4 presents the results of the exploratory factor rotation. The complete intercorrelation matrix of the 16 testlets appears in Appendix F.

Inspection of Table 4 indicates that quite good simple structure was obtained for the reading, mathematics, and HCG testlets. That is, the testlets marking a reading factor, mathematics factor, and an HCG factor tended to have large loadings only on their respective factors. The science testlets, however, appear to be somewhat more complex and show salient loadings on the reading and mathematics factors. That is, the chemistry testlet loaded on the mathematics factor as well as on the science factor. Similarly, the life science testlet loaded to a certain extent on the reading factor in addition to its more salient loading on the science factor. This does not come as a surprise since the internal consistency reliability of the Science test was lower than was the case for the other tests.

While the reading, mathematics, and HCG testlets demonstrated good convergent validity, the discriminant validity as measured by the factor inter-correlations was also reasonably encouraging. The correlation between reading and mathematics was .76 which approximates that found in typical factor analysis of the SAT. One might expect somewhat higher correlations between the NELS:88 verbal and mathematics factors than for their SAT counterparts since the NELS:88 sample is considerably less subject to selection than the SAT sample. Generally the factor correlations appear to vary little between the content areas and ranged from a low of .73 between Mathematics and History/Citizenship/ Geography and a high of .80 between History/Citizenship/ Geography and Science.

It is expected that the correlations among these factors will be somewhat reduced as the students begin to sort themselves out into various curriculum tracks as they go on to their last four years of high school.

Table 4.--Factor structure, NELS:88 tests

	PROMAX ROTATION			
TESTLETS	Factor 1	Factor 2	Factor 3	Factor 4
Read (literature)	.50	-.01	.08	.11
Read (science)	.39	.17	.03	.13
Read (poetry)	.62	.06	.00	.07
Read (biography)	.77	.00	.03	-.06
Read (history)	.64	.03	.02	-.02
Arithmetic	.02	.89	-.01	.02
Algebra	.08	.83	.03	-.06
Geometry	.00	.33	.02	.02
Probability	-.02	.44	.03	.11
Earth Science	.00	.05	.14	.59
Life Science	.21	.11	.04	.39
Chemistry	-.01	.29	.02	.39
Scientific Method	.21	.03	.02	.26
History	.04	-.01	.75	.05
Citizenship/Government	.11	.10	.63	-.02
Geography/Econ. Dev.	.11	.08	.37	.19

	FACTOR INTERCORRELATIONS			
	1	2	3	4
Factor 1	1.00			
Factor 2	.76	1.00		
Factor 3	.79	.73	1.00	
Factor 4	.75	.75	.80	1.00

SOURCE: U.S. Department of Education, National Center for Education Statistics, NELS:88 Base Year Survey.

Performance of Racial/Ethnic and Gender Groups on the NELS:88 Eighth Grade Test Battery

Table 5 presents means and standard deviations on the NELS:88 eighth grade tests by racial/ethnic and gender groups. These means are based on Item Response Theory (IRT) scoring using the three parameter IRT model (Lord & Novick, 1968) and the test weights. The scores used in these computations are the number right "true" scores corrected for guessing. The column in Table 5 labeled as "SD-DIF" presents the mean differences between the racial/ethnic subgroups and white majority group in terms of standard deviation units. Similarly the mean difference between male and female students on each of the tests is also presented in terms of standard deviation units.

Inspection of Table 5 suggests that the mean differences in terms of standard deviation units between the non-Asian racial/ethnic groups and the White majority group is about the same magnitude as that which was found for the 1980 tenth grade HS&B sample. The eighth grade female students are doing somewhat better than the male students at reading and about as well in mathematics. At the same time, females are doing somewhat less well than the male students in both science and history/citizenship/geography. It would appear that as early as the eighth grade, female students are beginning to fall behind in science.

Proficiency Level Subscores by Subgroups

In addition to providing scores for each of the four test content areas, behaviorally anchored proficiency level scores will also be reported in Reading and Mathematics. These proficiency level scores attempt to relate meaningful behaviors to various points on the total score scale. Three levels of **mathematics proficiency** and two levels of **reading proficiency** will be reported in addition to the usual normative scores for eighth graders. The three proficiency levels in mathematics form a hierarchical scale with each succeeding level characterized by increased complexity and where proficiency at a higher level implies proficiency at the lower levels. This Guttman scale property provides a limited amount of diagnostic information. The three mathematics proficiency levels define the following types of achievement:

- Level 1- Students who are proficient at this level are able to successfully carry out simple arithmetical operations on whole numbers.

- Level 2- Students who are proficient at this level have successfully mastered all the Level 1 tasks above as well as having mastered simple operations with decimals, fractions, and roots.

Table 5.--Weighted means and standard deviations of IRT scores on the NELS:88 tests, by racial/ethnic groups and sex

	TOTAL GROUP		WHITE		ASIAN			HISPANIC			BLACK			AMERICAN INDIAN		
	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	SD-DIF*	MEAN	S.D.	SD-DIF*	MEAN	S.D.	SD-DIF*	MEAN	S.D.	SD-DIF*
READING	10.3	6.0	11.4	5.9	10.8	6.2	-0.1	7.8	5.5	-0.6	7.1	5.3	-0.7	6.9	5.2	-0.7
MATHEMATICS	16.0	11.3	18.0	11.0	19.9	12.2	0.2	11.0	9.9	-0.6	8.9	9.1	-0.8	9.4	9.0	-0.8
SCIENCE	9.9	5.7	10.9	5.6	10.6	6.0	-0.1	7.5	5.0	-0.6	6.3	4.5	-0.8	6.5	4.9	-0.8
HIST/CIT/GEOG	15.1	7.6	16.4	7.2	16.1	8.2	0.0	11.6	7.7	-0.6	11.2	6.8	-0.7	10.5	7.2	-0.8

	MALE		FEMALE		
	MEAN	S.D.	MEAN	S.D.	SD-DIF*
READING	9.6	6.1	11.0	5.9	0.2
MATHEMATICS	16.1	11.5	15.9	11.1	0.0
SCIENCE	10.3	6.0	9.5	5.4	-0.1
HIST/CIT/GEOG	15.4	7.9	14.8	7.3	-0.1

	NUMBER OF CASES						
	WHITE	ASIAN	HISPANIC	BLACK	AM.IND.	MALE	FEMALE
READING	15,756	1,500	3,005	2,858	308	11,755	11,887
MATHEMATICS	15,753	1,495	2,996	2,860	307	11,750	11,878
SCIENCE	15,758	1,493	2,995	2,845	307	11,750	11,865
HIST/CIT/GEOG	15,693	1,487	2,981	2,842	308	11,692	11,832

* Difference between subgroup mean and reference group mean in terms of the total group standard deviation. An associated negative sign indicates that the reference group (Whites for racial/ethnic comparisons; males for sex comparisons) had a higher mean.

SOURCE: U.S. Department of Education, National Center for Education Statistics, NELS:88 Base Year Survey.

- Level 3- Students who are proficient at this level have mastered the two lower proficiency levels and are able to successfully solve simple problem solving tasks. Unlike levels 1 and 2 which require the rote application of rules, performance at this level requires conceptual understanding and/or the development of a solution strategy.

Mayer, Larkin, & Kadine (1984), also present a hierarchical model based on four knowledge structures. However, their model emphasizes a hierarchy of cognitive processing skills which are most appropriate for mathematics tests such as the SAT-M which almost entirely emphasizes problem solving skills. Their four model components are factual/linguistic, algorithmic, schematic, and strategic. The eighth grade proficiency level model suggested here follows more of a learning or curriculum sequencing model than either the Mayer et al. model or a similar cognitive processing model developed for the SAT-M by Rock and Johnson (1989). A major feature shared, however, by the eighth grade curriculum sequencing model and the models espoused by Mayer et al. and Rock et al. is that the components are assumed to be sequentially dependent during problem solving. That is, for successfully implementing a schema the problem solver should have mastered the requisite factual/linguistic knowledge necessary to read the problem.

In a primarily achievement oriented mathematics test such as the NELS eighth grade mathematics test, it was felt that the hierarchical dependencies should follow the typical learning or curriculum sequence. That is, mastery of simple operations on whole numbers is a necessary but not sufficient condition for mastery of simple operations on decimals and fractions etc. As NELS proceeds through the upper grades it is likely that there will be fewer individual differences on the simple declarative or algorithmic knowledge and more between-individual variability on the problem solving skills. Thus, proportionately greater emphasis can be put on the development of problem solving skills in the succeeding followups. This does not mean that the simple declarative knowledge and algorithmic procedures will be missing from the tenth grade followup. In fact the hierarchically ordered skills model as presented here is particularly appropriate for the multi-level testing procedure which is to be implemented at the tenth grade. Since the tenth grade multi-level forms are tailored to groups of students classified by their achievement levels (based on their eighth grade performance), the lower level forms will have a greater proportion of the simple algorithmic operations while the second and highest level forms will increasingly consist of items requiring conceptual understanding and production level problem solving skills. The hierarchical skill conception leads quite naturally to the multi-level testing model.

Two kinds of proficiency score interpretations are available. The first kind of interpretation is consistent with the typical usage in the criterion referenced literature (Glaser, 1963). It simply states whether or not a student is above or below a given threshold, e.g., Level 1 performance. A second interpretation has a more normative slant in that it gives the probability that a given student is proficient at a given level, say Level 1. Each student will have three mathematics proficiency probabilities-one for each

of the three mathematics levels. Changes in an individual's proficiency probabilities as he or she goes from the eighth to the tenth grade indicate where on the development growth curve that individual is making progress. For example, an individual who increases his problem solving skills between eighth and tenth grade will show changes in the probability of being proficient at Level 3, but show little or no change in his or her probabilities of Level 1 or Level 2 proficiency.

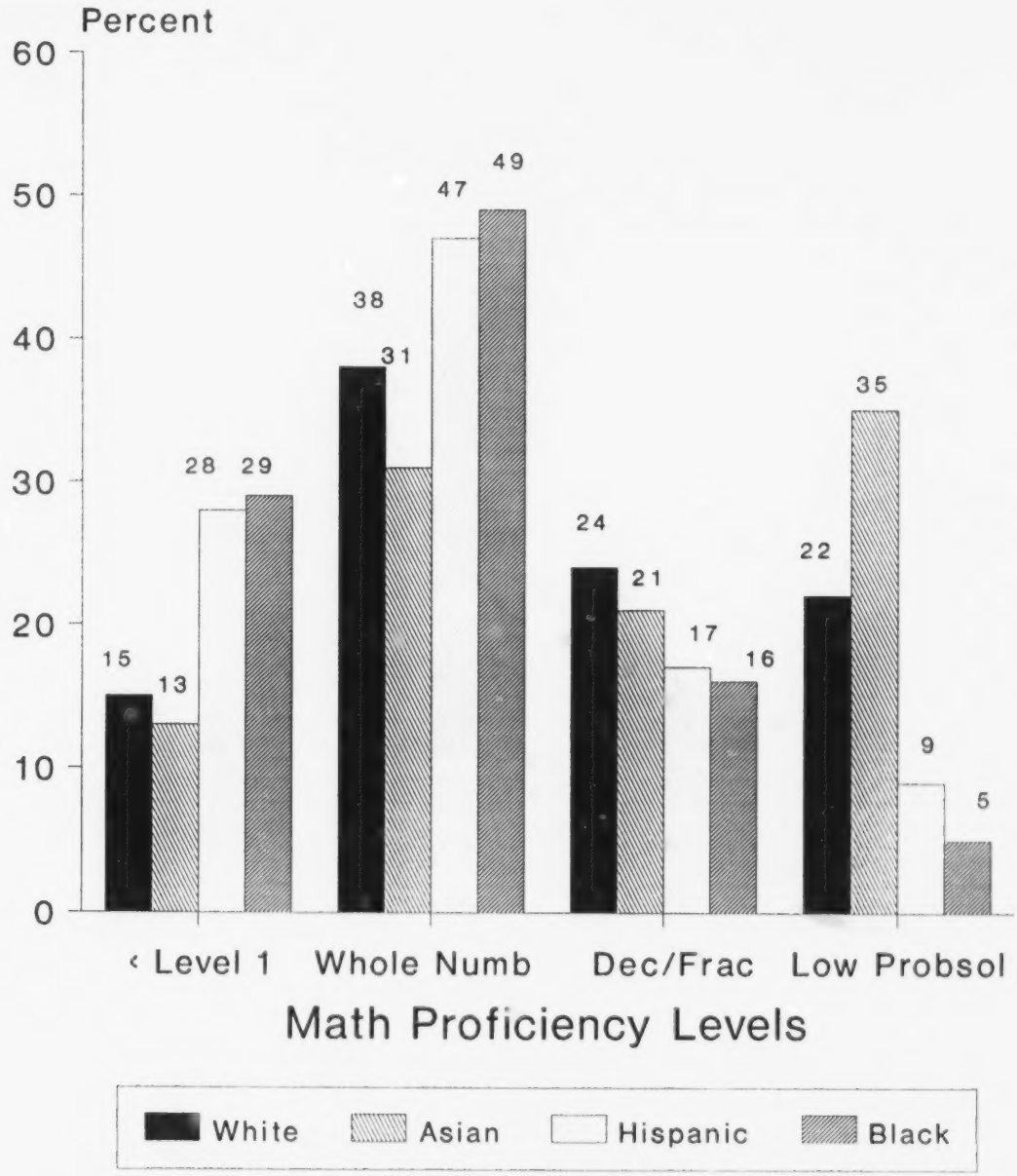
At this time, we will only present results on the criterion referenced type of interpretation. That is, we will report, for example, what percentage of a subgroup are proficient at Level 1 but have not mastered Level 2, and so on. Proficiency probabilities described in the second interpretation, which are most useful for measuring change over time, will be included in the presentation of results when grade 10 data are available.

Each proficiency level is marked by a block of 4 items that are relatively internally consistent with respect to the cognitive processes required. For example, level one marker items all deal with simple arithmetical operations on whole numbers. In addition to requiring the same cognitive operations, the items within a particular "marker" block should exhibit similar item difficulty parameters. Since the underlying cognitive demand model is assumed to be hierarchical, students who are proficient on the level 3 block of marker items should also demonstrate proficiency on the level 2 and level 1 items. If a student demonstrates proficiency on a higher level block but not on a lower level block, one must infer that the hierarchical model did not fit that particular individual. While four items may seem like a relatively small number of items, it should be remembered that all four are essentially parallel measures of the same content or processing skill. The four items are not a subscale that attempts to discriminate individuals all along a continuous dimension but are simply used to make a "go/no go" decision at a certain point referencing a specific skill. Evidence for the internal consistency of the hierarchical model is the low rate of reversals in the response patterns. About 95% of the students in all the subgroups had response patterns to the marker blocks that were consistent with the hierarchical model. See Appendix G for a detailed description of the way in which the proficiency scores were defined.

Figure 4 presents a proficiency profile of Racial/Ethnic groups on the mathematics test. It is clear from Figure 4 that there are relatively large group differences with respect to the type of problems that they can solve. Three-quarters (28% + 47%) of the eighth grade Hispanic students and nearly four-fifths (29% + 49%) of the Black students have not yet demonstrated proficiency with simple operations on decimals and fractions. Similarly, about 53% of the Whites and 44% of the Asians have yet to achieve proficiency in operations on decimals and fractions. The largest group differences occur at the most complex proficiency level which was defined by marker items requiring low level problem solving skills and/or conceptual understanding. The Asian students in particular are over represented at this proficiency level.

Figure 5 presents the mathematics proficiency profiles for the two sex groups. Inspection of Figure 5 indicates quite similar proficiency profile for the male and female students.

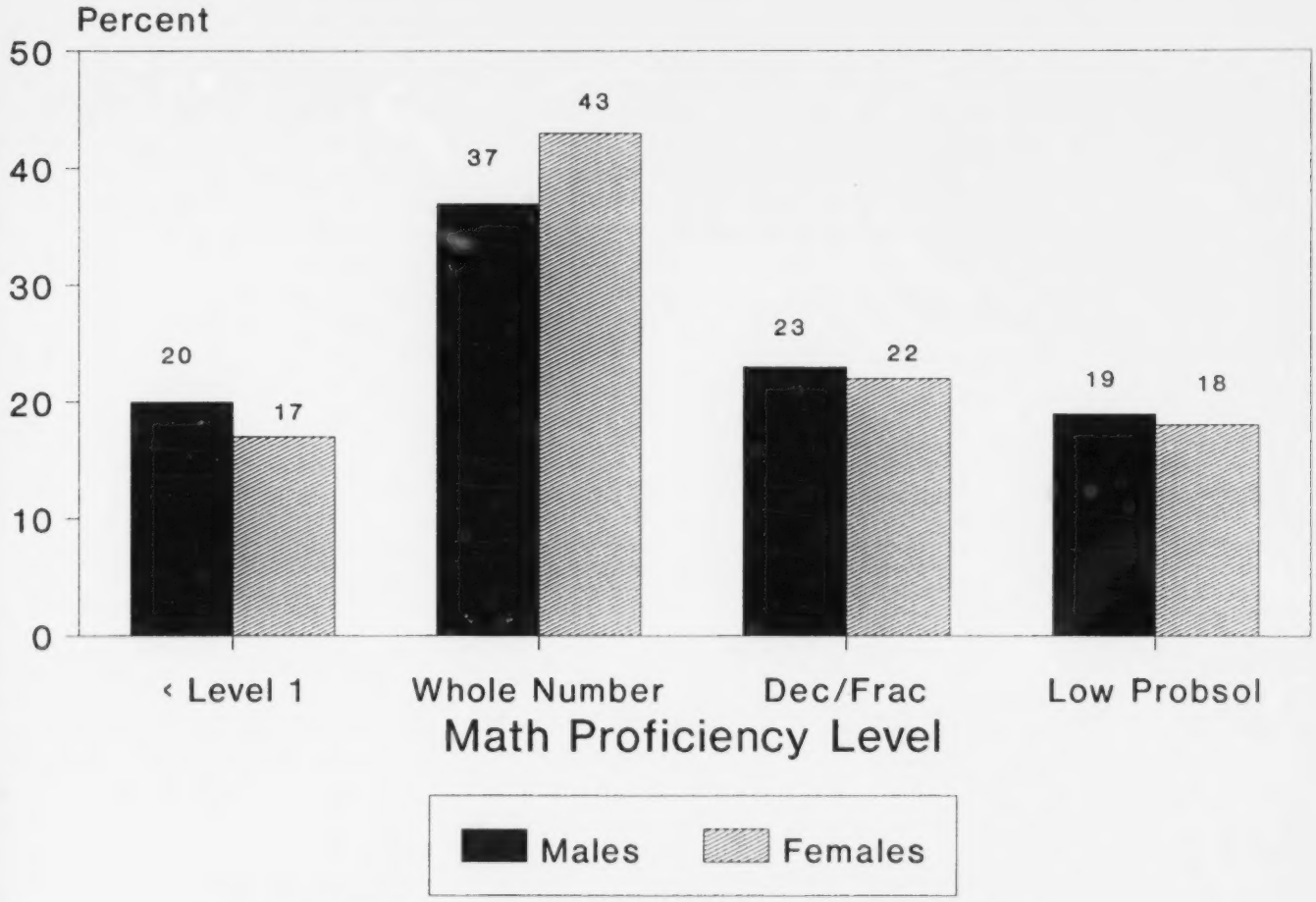
Figure 4.--Percent of selected subgroups that are proficient each mathematics proficiency level



SOURCE: U.S. Department of Education, National Center for Education Statistics, NELS:88: Base Year Survey.

BEST COPY AVAILABLE

Figure 5.--Percent of gender groups that are proficient at each mathematics proficiency level



SOURCE: U.S. Department of Education, National Center for Education Statistics, NELS:88:Base Year Survey

BEST COPY AVAILABLE

The two levels of proficiency that have been defined in the reading area are:

- Level 1- Simple reading comprehension including reproduction of detail and/or the author's main thought.

- Level 2- Ability to make inferences beyond the author's main thought and/or understand and evaluate relatively abstract concepts.

Figure 6 presents a reading level proficiency profile for selected racial/ethnic groups. As in the case of Mathematics, there are considerable differences between the groups with respect to the various mastery levels. The percentage of Asian and White students who have demonstrated proficiency at the inference level is about double that of the Hispanic and Black students.

Figure 7 presents the reading proficiency profile for the two sex groups. As in the case of mathematics, there is little difference between the patterns of proficiency for the sex groups at the eighth grade.

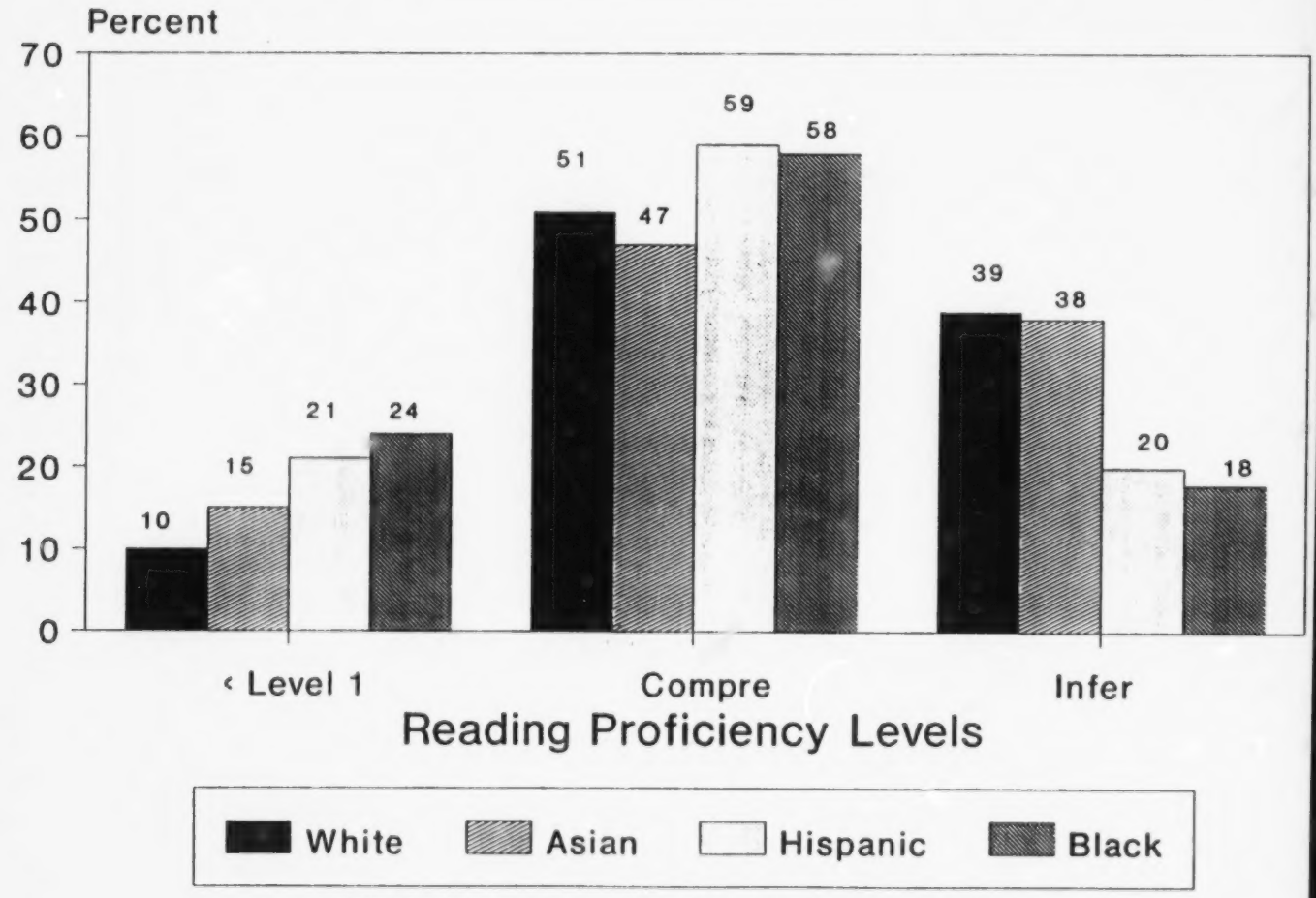
Item Response Theory (IRT) Parameters for the NELS:88 Battery

As pointed out above, the multi-stage testing strategy requires both vertical equating and lateral equating. That is, forms that vary between grade (vertical equating) as well as forms that vary within grade (lateral equating) must all be put on the same scale. The most efficient way of accomplishing this is to use Item Response Theory (IRT) equating. The previously reported item statistics (including the estimates of internal consistency reliability) support the feasibility of IRT scoring and eventually IRT based equating for at least the mathematics, reading, and History/Citizenship/ Geography tests. The following section provides further evidence of the relatively unifactorial nature of these three tests and thus their appropriateness for IRT applications.

Tetrachoric correlations among items within a content area were estimated and corrected for guessing. Principal components analysis was performed on each of the content area tetrachoric matrices. One simple factor analytic measure of the relative unidimensionality of the content areas is the ratio of the first and largest component to the second component (Reckase,1979; Hulin, Drasgow, & Parsons,1983). These ratios for reading, mathematics, science, and history/citizenship were 10:1, 12:1, 6:1, and 6:1. While all four show a single dominant factor, the reading and mathematics measures show a particularly dominant single factor. These results based on guessing-corrected tetrachoric matrices suggest that IRT estimation would provide reasonable estimates in all four content areas.

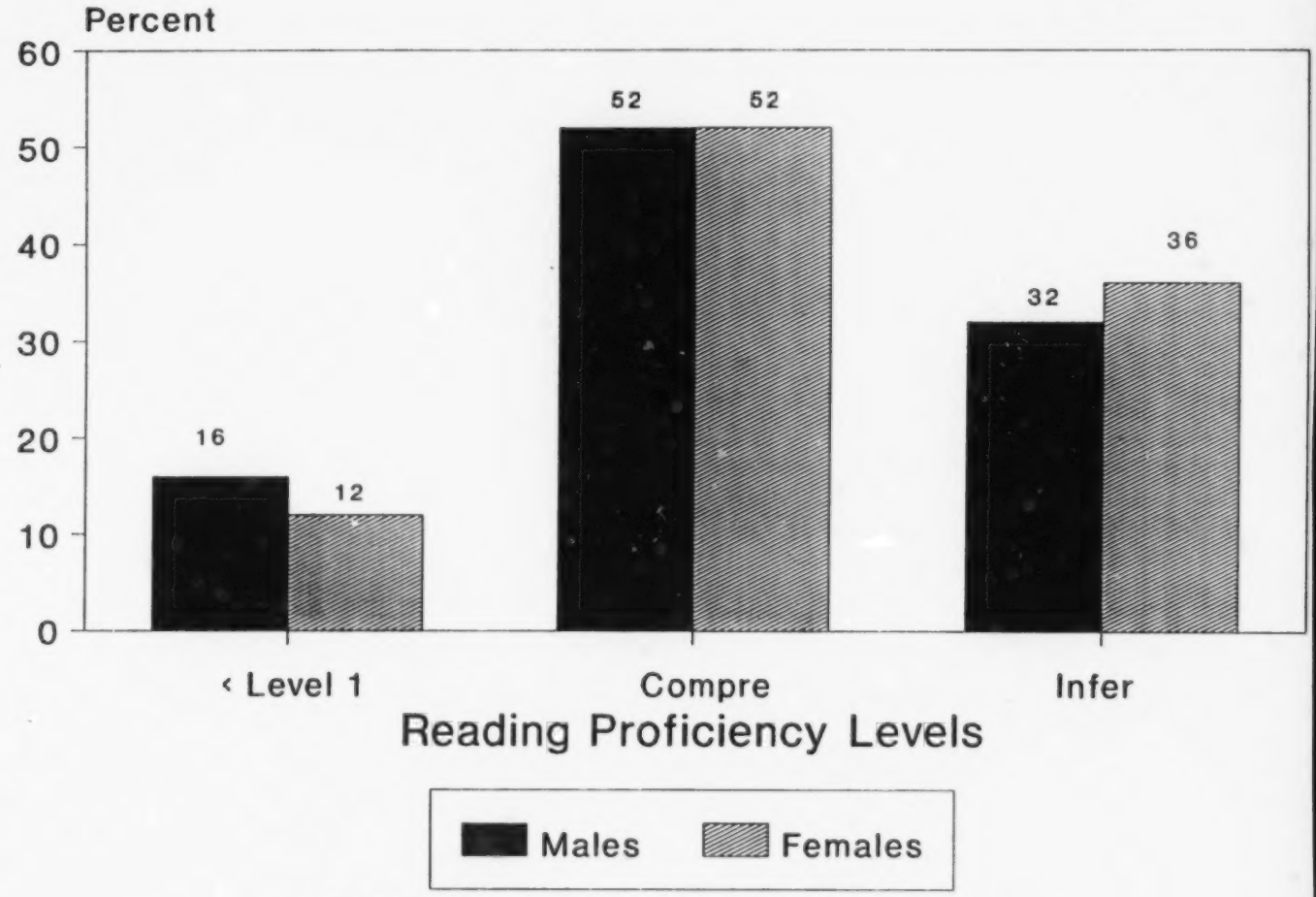
While factor analytic or principal component methods provide some useful information on the unidimensionality of the respective item pools, Lord often argued that one should go ahead and compute the IRT parameters and then examine the discrimination indices and the item trace lines for lack of fit. A monotonically

Figure 6.--Percent of selected subgroups that are proficient at each reading proficiency level



SOURCE: U.S. Department of Education, National Center for Education Statistics, NELS:88: Base Year Survey.

Figure 7.--Percent of gender groups that are proficient at each reading proficiency



SOURCE: U.S. Department of Education, National Center for Education Statistics, NELS:88: Base Year Survey

increasing trace line that comes close to the mean proportion correct for clusters of examinees grouped by ability level is evidence that the IRT model is a good description for the item and the test.

Appendices C1-C4 present the IRT item parameters for the reading, mathematics, science, and history/citizenship/geography eighth grade tests. The item parameters were computed using the Logist program (Wood et al., 1976). Item response theory (IRT) describes the probability of answering an item correctly as a mathematical function of ability level and characteristics of the items. The mathematical function used here, the logistic function, has one parameter for each individual's ability level and three parameters characterizing each item (Lord, 1980; Lord & Novick, 1968). The item parameters reflect difficulty level (b_i), discriminating power (a_i), and the likelihood of low ability individuals guessing the right answer (c_i). The function that relates the probability of passing a particular item i for a person of ability θ in terms of the item parameters is:

$$P_i(\theta) = c_i + (1 - c_i) \frac{1}{1 + \exp [- Da_i(\theta - b_i)]} \qquad (1)$$

where $D \doteq 1.7$

b_i = item difficulty, corresponding to the value of θ halfway between the guessing parameter and 1.0

a_i = discrimination parameter reflecting the steepness of the item characteristic curve at its point of inflection

c_i = "guessing parameter" probability of a person with very low ability getting the item correct

θ = a person's ability parameter usually standardized with mean 0 and standard deviation of 1.0

and $P_i(\theta)$ = probability of correct response of a person of ability level θ.

A person's number right true score (NRTS) is the simple sum of that particular person's $P_i(\theta)$'s. Thus, the scoring weights each item receives in the summation to arrive at NRTS are a function of the interaction of the item parameters with the person's θ or ability level. That is, the item characteristic functions, $P_i(\theta)$'s, provide a different score for a given item, depending upon a person's ability level. Inspection of the item characteristic function in equation (1) suggests that, for high ability people, the item score for a given item i will primarily depend on how much higher the person's θ is compared to the item difficulty (b_i, also measured in θ units), and how discriminating the item is.

A low-ability person will get little credit on a difficult item, even if he or she were to get it correct, because the model argues that the correct answer was probably guessed. This readily follows from equation (1). Such a person might have a θ (ability level) that was negative, say -1.5, and the b_i for a difficult item on the θ scale might be 2.0, and, since a_i is always positive, the denominator of equation (1) would become large in relation to the numerator. The limit here as the denominator gets larger is a scoring weight $P_i(\theta)$ equal to c_i the guessing parameter.

The fact that the item scores that are summed to get the number right true score are a function of the person's ability level θ, discrimination, difficulty, and guessing parameters, suggests that IRT scoring can be beneficial if (1) people with low ability can get the right answer by guessing; (2) items in the test vary in both difficulty and discrimination and thus an optimal scoring procedure should take this into account; (3) there are test center administration irregularities with respect to directions or timing that may lead to varying levels of items attempted and (4) the purpose is to put tests that share some but not all of the same items on the same scale.

Inspection of appendices C1-C4 indicate that only one item had a discrimination index ("a" parameter) in the thirties. This was a reading item (item 10) which had a difficulty parameter ("b") of 1.7, indicating that it was relatively difficult. The item was classified as requiring an inferential cognitive step. This item's biserial was in the forties (Appendix A1) suggesting that it may be reasonably reliable from the traditional psychometric viewpoint.

The summary statistics at the bottom of each column give the mean and standard deviation for each test's item parameters. In three out of four of the tests, the average discrimination parameter was greater than unity. In the 4th test, science, the average discrimination was only slightly less than unity (.98). Item discrimination parameters 1.0 and above are considered very good. Further investigation of the residuals for each item trace curve (not shown here) suggest that the IRT model fit quite well in reading, mathematics, history/citizenship/geography, and was reasonably acceptable in science.

With respect to both the skewness of the estimated theta distribution and the estimation of item parameters on the unweighted sample, Yamamoto (1990) has carried out empirical studies comparing weighted and unweighted, and skewed vs. unskewed theta distributions for both BILOG and LOGIST IRT estimation. His preliminary results suggest that there is bias in both the A and B parameters but LOGIST seems more robust when either the normality assumption is violated and/or the unweighted sample is used to estimate the IRT parameters. In spite of the fact that there may be differences in IRT parameters for various weightings/skewnesses, differences in theta means among various subgroups remain relatively invariant over violations of normality assumptions in the theta distributions and/or the use of weighted or unweighted samples. Work being carried out for NAEP may provide more information about this issue in the future.

Appendices D-1 through D-4 present test information functions for each of the tests. The information function is a simple transformation of the standard error of measurement: it is the reciprocal of the square of the SEM. Since it is impractical to present standard errors of measurement for each point in the score scale, the plot represents a picture of the estimated accuracy of measurement along the entire ability range. A high point on the plot corresponds to greater accuracy. For each of the four tests, the information function is above 1.0 for the ability range -2.0 to +2.0 (which includes more than 90% of the students), indicating a standard error of measurement of less than one score point in that range.

Test Scores on User Tape

The user tape of NELS:88 base year data available from NCES contains a variety of formulations of the test scores for the convenience of analysts. For each of the four cognitive tests, number of correct answers, number of wrong answers, and number of items omitted are included. A formula score for each test consists of the number right minus a proportion of the number wrong, and represents an effort to correct for score differences that are attributable to different response styles with respect to guessing, rather than to differences in knowledge of the correct answers. That is, one student may have a tendency to guess at random if he or she does not know the answer to a question, while another will simply leave the item blank. For four-choice test items, the expectation is that one fourth of the random guesses are likely to be correct, thus raising the number-right score for the student who chooses to guess over that of a student of equal ability who omits unknown items. The guessing correction subtracts a proportion of the wrong answers from the number right, with the proportion depending on the number of answer-choices for the items. In the case of four-choice items, again, the assumption is made that random guessing will produce approximately one-fourth correct answers and three-fourths wrong. So subtracting one-third of the incorrect answers from the number right produces an estimate of the score that would have been attained by another student of equal ability who chose to omit items instead of guessing. Computation of formula scores on the user tape took into account the number of answer choices for each incorrect item, that is, by subtracting $1/(n-1)$ for each wrong answer, where n is the number of response options. Omitted items are not treated as wrong, and do not enter into computation of formula scores.

IRT number-right scores, as discussed in detail in the section on IRT earlier, represent the sum of the probabilities of correct answers on each of the items in the test, given an individual's overall ability level. The IRT formula score on the user tape is a transformation of this score, in which a correction is made for the probability of an incorrect response, $1-P_i$, on each item. The correction factor, $(1-P_i)/(n-1)$ for each item, is subtracted from the IRT number-right score. While this is not necessary as a correction for guessing, since the possibility of guessing is already compensated for in the IRT model, the IRT formula score is preferred by some researchers since it more nearly approximates the range, mean, and variance of the raw formula score metric.

The final scores included in the NELS:88 user tape are standardized scores for each test, with each content area scaled to an estimated national mean of 50 and standard deviation of 10. This is accomplished by simply subtracting the weighted overall mean from each raw formula score, dividing by the standard deviation, multiplying by 10, and adding 50. Analysts find this formulation useful because it provides a convenient framework for comparison of individual or subgroup scores with national averages. For example, a subgroup average of 55 in standardized units represents an achievement level half a standard deviation higher than the national average. The standardized composite on the user tape is the average of the reading and mathematics standardized scores.

Quartile scores based on the raw formula score for each content area, as well as for the standardized composite, are included on the tape. These simply break each weighted score distribution into fourths, and are included for the convenience of users who require a simple way of dividing the sample by achievement level.

Approximately 4% of the 24,599 students who completed questionnaires did not have test scores. There were several reasons for missing test scores: (1) In some cases, initial parent refusal to let the student participate was turned around when the parent was recontacted for the parent survey in the summer. In such cases, students were interviewed by telephone, but no tests were administered. (2) Several schools refused the test component of the survey because of the time burden but agreed to do the student questionnaire. (3) In school-administered makeup days, typically only the student questionnaire was administered. (4) Some materials were lost in transit. In some of these cases the questionnaire was then administered by telephone, but not the test. (5) Some of the students were present for the test administration but failed to answer items in one or more sections of the test. Test sections were not scored if fewer than five items were answered. Special sample weights adjusted for test nonresponse were used for analyses in this report, and differ in this respect from the basic student weight (BYQWT) on the public use tape.

CHAPTER 4. CONCLUSIONS

The results suggest that for the most part the NELS:88 eighth grade test battery either met or exceeded its psychometric objectives. While the allotted testing time was only about one and a half hours, quite acceptable reliabilities were obtained for the Reading Comprehension, Mathematics, and the History/Citizenship/Geography test. In fact, the NELS:88 battery reliabilities significantly exceeded their counterparts in the previous HS&B test battery.

These internal consistency reliabilities were sufficiently high to justify the use of Item Response Theory (IRT) scoring, and thus provide the framework for constructing follow-up forms that will be more adaptive to the ability level of the student. The IRT scaling will enable the researcher to administer forms varying in difficulty (at the tenth grade) depending on the student's previous (eighth grade) achievement scores in the areas of Reading, Mathematics, and possibly History/Citizenship/Geography. This adaptive approach will both minimize potential ceiling effects when the students are followed up as tenth graders, and it will also help to increase measurement accuracy.

The Science test was considerably less unifactorial than the other tests. This finding poses less of a problem in the Science area since there appears to be little possibility of ceiling effects at least up to and including the tenth grade. Thus, there appears to be little need for a tenth grade form that is adaptive.

There was little evidence of differential item functioning (DIF) for either gender or racial/ethnic groups.

Factor analytic results supported the discriminant validity of the four content areas. Convergent validity was also indicated by the salient loadings of the testlets composed of "marker items" on their hypothesized factors.

In addition to providing the usual normative scores in all four tested areas, behaviorally anchored proficiency level scores are available in both the Reading and Mathematics areas on the NELS:88 public release tapes.

REFERENCES

Airasian, P. W., & Madaus, G. F. (1983). Linking testing and instruction: Policy issues. Journal of Educational Measurement, 20, 103-118.

Bock, R. D., Gibbons, R., & Muraki, E. (1985). Full-information item factor analysis (MRC Report No. 85-1). Chicago: National Opinion Research Center, University of Chicago.

Cole, N. S., & Moss, P. A. (1989). Bias in Test Use. In Robert L. Linn (Ed.) Educational Measurement, 3rd edition. National Council on Measurement in Education and American Council on Education. New York: MacMillan Publishing Company.

Cole, N. S., & Nitko, A. J. (1981). Instrumentation and bias: Issues in selecting measures for educational evaluations. In R. A. Berk (Ed.), Educational evaluation methodology: The state of the art. Baltimore: Johns Hopkins University Press.

Cooley, W. W. (1977, August). Program evaluation in education. Invited paper presented at the meeting of the American Psychological Association, San Francisco.

Cronbach, L. J. (1963). Evaluation for course improvement. Teachers College Record, 64, 672-683

Frechtling, J. A. (1989). Administrative uses of school testing programs. In R. L. Linn (ed.) Educational Measurement 3rd Edit., 475-483. New York: MacMillan.

Glaser, R. (1963). Instructional technology and the measurement of learning outcomes: some questions. American Psychologist, 18, 519-521.

Guttman, L. (1954). A new approach to factor analysis. The radex. In P. F. Lazarfield (Ed), Mathematical thinking in the social sciences, 216-257. New York: Free Press.

Hendricksen, A. E. & White, P. O. (1964). Promax: A quick method for rotation to oblique simple structure. British Journal of Statistical Psychology, 17, 65-70.

Holland, P. W., & Thayer, D. T. (1986). Differential item functioning and the Mantel-Haenszel procedure. ETS Research Report No. 86-31. Princeton, NJ.

Hulin, C. L., Drasgow, F., & Parsons, C.K. (1983). Item response theory. Homewood: Dow-Jones Irwin.

Ingels, S. et al. (1990). Base Year Student Component Data User's Manual, National Educational Longitudinal Study of 1988 (NELS:88). U.S. Government Printing Office, District of Columbia.

Leinhardt, G. (1980). Modeling and measuring educational treatment in evaluation. Review of Educational Research, 50, 393-420.

Leinhardt, G., & Seewald, A. M. (1981). Overlap: What's tested, what's taught? Journal of Educational Measurement, 18, 85-96.

Linn, R. L., & Harnisch, D. L. (1981). Interactions between item content and group membership on achievement test items. Journal of Educational Measurement, 18, 109-118.

Lord, F. M. (1980). Applications of item response theory to practical testing problems. Hillsdale, N. J.: Erlbaum.

Lord, F. M. & Novick, M. R. (1968). Statistical theories of mental test scores. Reading Mass.: Addison-Wesley.

Mayer, R., Larkin, J., & Kadane, P. (1984). A cognitive analysis of mathematical problem solving ability. In R. J. Sternberg (Ed.) Advances in the psychology of human intelligence, Vol. 2, pp. 231-273. Hillsdale, NJ: Lawrence Erlbaum Associates.

Mehrens, W. A. (1984). National tests and local curriculum: Match or mismatch? Educational Measurement: Issues and Practice, 3 (3), 9-15.

Mehrens, W. A., & Phillips, S. E. (1986). Detecting impacts of curricular differences in achievement test data. Journal of Educational Measurement, 23, 185-196.

Nitko, A. J. (1983). Educational tests and measurements: An introduction. San Diego: Harcourt, Brace Jovanovich.

Nitko, A. J. (1989). Designing tests that are integrated with instruction. In Robert L. Linn (Ed.) Educational Measurement, 3rd edition. National Council on Measurement in Education and American Council on Education. New York: MacMillan Publishing Company.

Phillips, S. E., & Mehrens, W. A. (1988). Effects of curricular differences on achievement test data at item and objective levels. Applied Measurement in Education, 1, 33-51.

Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. Journal of Educational Statistics, 4, 207-230.

Rock, D. A., Hilton T. F., Pollack, J., Ekstrom, R. & Goertz, M. (1985). Psychometric analysis of the NLS and the High School and Beyond Test Batteries: Contractors Report #CES85-218. U.S. Govt. Printing Office, District of Columbia.

Rock, D. A., & Johnson, E. (1990). Criterion referenced scoring of the SAT-M. Submitted for review. College Board Report.

Rock, D. & Pollack (1987) Development of the NELS:88 Cognitive Test Battery in Ingels et al (1990) NELS:88 Base Year Field Test Report. Chicago Ill. NORC, ERIC document ED 289-897.

Schmidt, W. H. (1983). Content bias in achievement tests. Journal of Educational Measurement, 20, 165-178.

Spencer, B.D. et al (1990) NELS:88 Base Year Sample Design Report. Washington, DC, NCES.

Wood, R. L., Wingersky, M. S., & Lord, F. M. (1976). LOGIST: A computer program for estimating examinee ability and item characteristic curve parameters. ETS, Research Memorandum #76-6. Princeton, NJ.

Yamamoto, K., personal communication, 1990.

APPENDIX A

ITEM ANALYSIS STATISTICS

Appendix A-1

Item Analysis Statistics, Reading

	TOTAL			MALE			FEMALE		
	P+	RBIS	DELTA	P+	RBIS	DELTA	P+	RBIS	DELTA
ITEM 1	0.95	0.59	6.5	0.93	0.60	7.0	0.96	0.56	5.9
ITEM 2	0.85	0.62	8.8	0.85	0.61	8.9	0.86	0.64	8.7
ITEM 3	0.82	0.65	9.3	0.80	0.63	9.7	0.85	0.67	8.9
ITEM 4	0.57	0.66	12.3	0.53	0.65	12.7	0.62	0.66	11.8
ITEM 5	0.55	0.67	12.5	0.53	0.62	12.7	0.57	0.71	12.3
ITEM 6	0.60	0.65	12.0	0.61	0.68	11.9	0.60	0.63	12.0
ITEM 7	0.41	0.63	13.9	0.39	0.64	14.1	0.42	0.62	13.8
ITEM 8	0.49	0.68	13.1	0.48	0.66	13.2	0.50	0.70	13.0
ITEM 9	0.61	0.56	11.9	0.56	0.55	12.4	0.66	0.57	11.3
ITEM 10	0.39	0.45	14.1	0.38	0.50	14.2	0.40	0.39	14.0
ITEM 11	0.59	0.65	12.1	0.54	0.65	12.6	0.63	0.63	11.6
ITEM 12	0.71	0.76	10.8	0.66	0.75	11.4	0.76	0.75	10.2
ITEM 13	0.50	0.55	13.0	0.52	0.56	12.8	0.49	0.56	13.1
ITEM 14	0.48	0.65	13.2	0.45	0.64	13.5	0.50	0.65	13.0
ITEM 15	0.46	0.70	13.4	0.43	0.70	13.7	0.49	0.70	13.1
ITEM 16	0.76	0.74	10.1	0.73	0.75	10.5	0.79	0.73	9.8
ITEM 17	0.53	0.67	12.7	0.49	0.64	13.1	0.57	0.69	12.3
ITEM 18	0.54	0.53	12.6	0.51	0.51	12.9	0.56	0.55	12.4
ITEM 19	0.63	0.68	11.7	0.59	0.65	12.0	0.66	0.70	11.4
ITEM 20	0.70	0.64	10.9	0.67	0.63	11.3	0.74	0.65	10.4
ITEM 21	0.62	0.62	11.8	0.60	0.59	12.0	0.64	0.65	11.5
COLUMN MEAN	0.61	0.64	11.7	0.58	0.63	12.0	0.63	0.64	11.4
COLUMN S.D.	0.14	0.07	1.8	0.14	0.06	1.7	0.15	0.08	1.9

SAMPLE SIZE	23679	11689	11814
POPULATION ESTIMATE	3005290	1495064	1491180
COEFFICIENT ALPHA	0.84	0.84	0.83
SPLIT HALF RELIABILITY	0.85	0.85	0.85

	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.
FORMULA SCORE	10.2	6.16	9.5	6.21	10.9	6.03
NUMBER RIGHT	12.6	4.81	12.1	4.85	13.2	4.70
NUMBER WRONG	8.0	4.64	8.4	4.68	7.5	4.54
NUMBER OMITS	0.2	0.65	0.2	0.69	0.2	0.61
NUMBER NOT REACHED	0.2	1.26	0.3	1.42	0.2	1.07

Source: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988: Base Year Survey .

Appendix A-1--(continued)

Item Analysis Statistics, Reading

	TOTAL			ASIAN			HISPANIC			BLACK			WHITE			AMERICAN INDIAN		
	P+	RBIS	DELTA	P+	RBIS	DELTA	P+	RBIS	DELTA	P+	RBIS	DELTA	P+	RBIS	DELTA	P+	RBIS	DELTA
ITEM 1	0.95	0.59	6.5	0.95	0.70	6.6	0.93	0.54	7.2	0.93	0.49	7.1	0.95	0.63	6.2	0.95	0.35	6.4
ITEM 2	0.85	0.62	8.8	0.85	0.66	8.9	0.80	0.58	9.7	0.75	0.55	10.2	0.88	0.62	8.2	0.72	0.53	10.7
ITEM 3	0.82	0.65	9.3	0.80	0.70	9.6	0.75	0.61	10.4	0.73	0.58	10.5	0.85	0.64	8.8	0.72	0.67	10.7
ITEM 4	0.57	0.66	12.3	0.56	0.62	12.4	0.46	0.64	13.4	0.38	0.62	14.2	0.63	0.64	11.7	0.45	0.59	13.5
ITEM 5	0.55	0.67	12.5	0.54	0.69	12.5	0.41	0.63	13.9	0.35	0.60	13.6	0.59	0.66	12.0	0.36	0.61	14.4
ITEM 6	0.60	0.65	12.0	0.63	0.71	11.7	0.49	0.61	13.1	0.44	0.55	13.6	0.65	0.64	11.4	0.45	0.68	13.5
ITEM 7	0.41	0.63	13.9	0.43	0.69	13.7	0.29	0.55	15.2	0.26	0.52	15.6	0.45	0.62	13.5	0.26	0.59	15.6
ITEM 8	0.49	0.68	13.1	0.54	0.71	12.6	0.36	0.66	14.4	0.35	0.62	14.5	0.54	0.66	12.6	0.33	0.76	14.8
ITEM 9	0.61	0.56	11.9	0.66	0.51	11.3	0.55	0.54	12.5	0.51	0.53	12.9	0.64	0.57	11.6	0.50	0.42	13.0
ITEM 10	0.39	0.45	14.1	0.43	0.45	13.7	0.34	0.45	14.6	0.32	0.40	14.9	0.42	0.44	13.8	0.29	0.51	15.2
ITEM 11	0.59	0.65	12.1	0.64	0.64	11.6	0.54	0.55	12.6	0.46	0.56	13.4	0.62	0.66	11.8	0.48	0.53	13.2
ITEM 12	0.71	0.76	10.8	0.70	0.77	10.9	0.61	0.68	11.9	0.52	0.66	12.8	0.76	0.76	10.2	0.56	0.73	12.4
ITEM 13	0.50	0.55	13.0	0.54	0.62	12.6	0.43	0.44	13.7	0.38	0.38	14.2	0.54	0.58	12.6	0.35	0.35	14.6
ITEM 14	0.48	0.65	13.2	0.52	0.70	12.8	0.37	0.53	14.3	0.37	0.54	14.3	0.51	0.67	12.9	0.34	0.52	14.6
ITEM 15	0.46	0.70	13.4	0.51	0.72	12.9	0.36	0.64	14.4	0.36	0.69	14.5	0.50	0.70	13.0	0.34	0.62	14.6
ITEM 16	0.76	0.74	10.1	0.79	0.71	9.8	0.67	0.66	11.3	0.65	0.66	11.4	0.80	0.76	9.6	0.60	0.70	11.9
ITEM 17	0.53	0.67	12.7	0.57	0.64	12.3	0.39	0.54	14.2	0.40	0.49	14.0	0.58	0.69	12.2	0.42	0.41	13.9
ITEM 18	0.54	0.53	12.6	0.56	0.51	12.4	0.48	0.47	13.2	0.45	0.52	13.5	0.56	0.53	12.4	0.36	0.54	14.5
ITEM 19	0.63	0.68	11.7	0.65	0.69	11.4	0.52	0.56	12.8	0.45	0.58	13.5	0.67	0.68	11.2	0.46	0.53	13.4
ITEM 20	0.70	0.64	10.9	0.74	0.63	10.5	0.63	0.57	11.7	0.57	0.55	12.3	0.74	0.66	10.4	0.59	0.56	12.0
ITEM 21	0.62	0.62	11.6	0.62	0.60	11.7	0.50	0.53	13.0	0.48	0.48	13.2	0.67	0.63	11.3	0.47	0.51	13.3
COLUMN MEAN	0.61	0.64	11.7	0.63	0.65	11.5	0.52	0.57	12.7	0.49	0.55	13.0	0.65	0.64	11.3	0.48	0.56	13.2
COLUMN S.D.	0.14	0.07	1.8	0.13	0.08	1.7	0.16	0.07	1.9	0.16	0.08	1.9	0.14	0.07	1.8	0.16	0.11	2.0

SAMPLE SIZE	23679	1500	3003	2871	15771	308
POPULATION ESTIMATE	3005290	105759	304711	391769	2129481	43293
COEFFICIENT ALPHA	0.84	0.85	0.79	0.77	0.83	0.78
SPLIT HALF RELIABILITY	0.85	0.87	0.81	0.80	0.84	0.78

	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.
FORMULA SCORE	10.2	6.16	10.8	6.28	7.7	5.63	6.9	5.43	11.3	6.00	6.7	5.52
NUMBER RIGHT	12.6	4.81	13.1	4.91	10.7	4.44	10.0	4.28	13.5	4.65	9.9	4.34
NUMBER WRONG	8.0	4.64	7.5	4.74	9.7	4.26	10.2	4.26	7.2	4.53	10.5	4.24
NUMBER OMITS	0.2	0.65	0.2	0.57	0.2	0.76	0.3	0.83	0.2	0.58	0.4	1.00
NUMBER NOT REACHED	0.2	1.26	0.2	1.36	0.4	1.68	0.6	2.03	0.1	0.90	0.3	1.29

Source: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988: Base Year Survey .

Appendix A-1--(continued)

Item Analysis Statistics, Reading

	HISPANIC MALE			HISPANIC FEMALE			BLACK MALE			BLACK FEMALE			WHITE MALE			WHITE FEMALE		
	P+	RBIS	DELTA	P+	RBIS	DELTA	P+	RBIS	DELTA	P+	RBIS	DELTA	P+	RBIS	DELTA	P+	RBIS	DELTA
ITEM 1	0.92	0.56	7.4	0.94	0.52	6.8	0.91	0.48	7.6	0.95	0.50	6.4	0.94	0.65	6.7	0.97	0.58	5.6
ITEM 2	0.79	0.57	9.8	0.80	0.58	9.6	0.75	0.54	10.4	0.77	0.56	10.1	0.88	0.61	8.3	0.89	0.63	8.1
ITEM 3	0.72	0.58	10.7	0.77	0.64	10.0	0.71	0.57	10.7	0.75	0.58	10.3	0.83	0.62	9.2	0.88	0.66	8.4
ITEM 4	0.42	0.63	13.8	0.49	0.64	13.1	0.34	0.61	14.6	0.42	0.62	13.8	0.59	0.64	12.1	0.67	0.64	11.2
ITEM 5	0.41	0.59	13.9	0.41	0.67	13.9	0.43	0.56	13.7	0.46	0.64	13.4	0.57	0.62	12.3	0.62	0.70	11.8
ITEM 6	0.50	0.64	13.0	0.48	0.58	13.2	0.45	0.57	13.5	0.44	0.54	13.6	0.66	0.67	11.3	0.65	0.62	11.5
ITEM 7	0.28	0.58	15.4	0.30	0.53	15.1	0.26	0.50	15.5	0.26	0.56	15.6	0.44	0.64	13.6	0.47	0.60	13.3
ITEM 8	0.36	0.66	14.4	0.37	0.67	14.4	0.35	0.59	14.6	0.36	0.66	14.4	0.52	0.65	12.8	0.55	0.68	12.5
ITEM 9	0.51	0.53	12.9	0.58	0.54	12.2	0.45	0.48	13.5	0.57	0.57	12.3	0.58	0.55	12.2	0.69	0.57	11.0
ITEM 10	0.33	0.53	14.7	0.35	0.37	14.5	0.30	0.48	15.1	0.34	0.32	14.7	0.41	0.49	13.9	0.42	0.39	13.8
ITEM 11	0.52	0.52	12.8	0.55	0.58	12.4	0.41	0.60	13.9	0.51	0.52	12.9	0.57	0.68	12.3	0.67	0.64	11.3
ITEM 12	0.56	0.66	12.4	0.66	0.69	11.3	0.47	0.68	13.3	0.57	0.65	12.3	0.71	0.76	10.8	0.81	0.75	9.4
ITEM 13	0.45	0.46	13.5	0.41	0.43	13.9	0.41	0.43	13.9	0.36	0.35	14.4	0.55	0.58	12.5	0.53	0.59	12.7
ITEM 14	0.37	0.53	14.3	0.38	0.53	14.3	0.34	0.54	14.6	0.40	0.53	14.0	0.48	0.66	13.2	0.54	0.67	12.6
ITEM 15	0.34	0.66	14.7	0.34	0.62	14.2	0.33	0.71	14.7	0.38	0.67	14.2	0.46	0.70	13.4	0.53	0.70	12.7
ITEM 16	0.67	0.66	11.2	0.66	0.67	11.3	0.62	0.68	11.8	0.68	0.63	11.1	0.77	0.77	10.1	0.83	0.74	9.1
ITEM 17	0.37	0.54	14.3	0.40	0.53	14.0	0.36	0.44	14.4	0.44	0.54	13.6	0.54	0.66	12.6	0.62	0.72	11.8
ITEM 18	0.46	0.42	13.4	0.50	0.51	13.0	0.43	0.51	13.7	0.46	0.52	13.4	0.54	0.51	12.6	0.59	0.55	12.1
ITEM 19	0.51	0.54	12.9	0.52	0.58	12.8	0.43	0.52	13.7	0.47	0.64	13.3	0.64	0.67	11.6	0.71	0.69	10.7
ITEM 20	0.58	0.59	12.2	0.68	0.55	11.2	0.54	0.53	12.6	0.61	0.56	11.9	0.70	0.64	10.9	0.78	0.66	9.9
ITEM 21	0.50	0.51	13.0	0.50	0.56	13.0	0.46	0.40	13.4	0.51	0.54	12.9	0.64	0.61	11.5	0.69	0.65	11.0
COLUMN MEAN	0.50	0.57	12.9	0.53	0.57	12.6	0.46	0.54	13.3	0.51	0.56	12.8	0.62	0.64	11.6	0.67	0.64	11.0
COLUMN S.D.	0.16	0.07	1.8	0.16	0.08	1.9	0.16	0.08	1.8	0.16	0.09	2.0	0.14	0.07	1.8	0.14	0.08	1.9
SAMPLE SIZE	1437			1545			1386			1466			7831			7827		
POPULATION ESTIMATE	151316			151394			191961			197273			1061031			1055784		
COEFFICIENT ALPHA	0.79			0.79			0.76			0.78			0.84			0.83		
SPLIT HALF RELIABILITY	0.80			0.81			0.79			0.80			0.85			0.84		
	MEAN	S.D.		MEAN	S.D.		MEAN	S.D.		MEAN	S.D.		MEAN	S.D.		MEAN	S.D.	
FORMULA SCORE	7.3	5.61		8.1	5.61		6.2	5.31		7.5	5.48		10.5	6.12		12.0	5.78	
NUMBER RIGHT	10.4	4.43		11.0	4.42		9.4	4.21		10.5	4.29		12.9	4.75		14.1	4.47	
NUMBER WRONG	9.9	4.25		9.5	4.24		10.5	4.24		9.8	4.26		7.8	4.63		6.7	4.37	
NUMBER OMITS	0.3	0.79		0.2	0.73		0.3	0.79		0.3	0.84		0.2	0.63		0.1	0.53	
NUMBER NOT REACHED	0.4	1.75		0.4	1.62		0.8	2.37		0.4	1.63		0.2	1.02		0.1	0.75	

Source: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988: Base Year Survey .

Appendix A-2

Item Analysis Statistics, Mathematics

	TOTAL			MALE			FEMALE		
	P+	RBIS	DELTA	P+	RBIS	DELTA	P+	RBIS	DELTA
ITEM 1	0.71	0.69	10.8	0.69	0.69	11.8	0.72	0.69	10.6
ITEM 2	0.58	0.60	13.0	0.51	0.59	12.9	0.49	0.60	13.1
ITEM 3	0.47	0.27	13.3	0.46	0.28	13.4	0.49	0.25	13.1
ITEM 4	0.49	0.65	13.1	0.48	0.68	13.2	0.49	0.61	13.1
ITEM 5	0.51	0.65	12.9	0.51	0.65	12.9	0.50	0.65	13.0
ITEM 6	0.45	0.45	13.5	0.45	0.43	13.5	0.45	0.46	13.5
ITEM 7	0.41	0.69	13.9	0.41	0.68	14.0	0.42	0.70	13.8
ITEM 8	0.37	0.59	14.4	0.36	0.56	14.4	0.37	0.62	14.3
ITEM 9	0.44	0.51	13.6	0.46	0.50	13.4	0.42	0.52	13.9
ITEM 10	0.41	0.60	13.9	0.40	0.64	14.0	0.41	0.56	13.9
ITEM 11	0.35	0.54	14.5	0.37	0.52	14.3	0.33	0.56	14.7
ITEM 12	0.44	0.64	13.6	0.42	0.64	13.9	0.46	0.67	13.4
ITEM 13	0.52	0.70	12.8	0.54	0.70	12.6	0.51	0.69	12.9
ITEM 14	0.58	0.69	13.0	0.52	0.68	12.8	0.47	0.69	13.3
ITEM 15	0.71	0.50	10.8	0.69	0.52	11.0	0.72	0.49	10.7
ITEM 16	0.79	0.49	9.8	0.79	0.54	9.8	0.79	0.44	9.8
ITEM 17	0.70	0.46	10.9	0.70	0.49	10.9	0.70	0.42	10.9
ITEM 18	0.52	0.64	12.8	0.51	0.63	12.9	0.52	0.65	12.8
ITEM 19	0.79	0.58	9.7	0.78	0.60	10.0	0.81	0.56	9.5
ITEM 20	0.79	0.59	9.8	0.76	0.55	10.2	0.82	0.45	9.3
ITEM 21	0.69	0.55	11.0	0.73	0.59	10.6	0.65	0.53	11.5
ITEM 22	0.68	0.71	11.1	0.70	0.71	10.9	0.67	0.71	11.3
ITEM 23	0.65	0.45	11.5	0.65	0.44	11.4	0.64	0.46	11.5
ITEM 24	0.59	0.57	12.1	0.59	0.57	12.1	0.59	0.56	12.1
ITEM 25	0.65	0.65	11.4	0.71	0.62	10.8	0.59	0.69	12.0
ITEM 26	0.62	0.59	11.7	0.64	0.60	11.6	0.61	0.59	11.9
ITEM 27	0.60	0.76	12.0	0.62	0.75	11.8	0.59	0.77	12.1
ITEM 28	0.56	0.59	12.4	0.54	0.62	12.6	0.59	0.57	12.1
ITEM 29	0.52	0.64	12.8	0.50	0.68	13.0	0.53	0.64	12.7
ITEM 30	0.52	0.54	12.8	0.50	0.56	13.0	0.54	0.51	12.6
ITEM 31	0.59	0.67	12.1	0.61	0.67	11.9	0.58	0.67	12.2
ITEM 32	0.66	0.54	11.3	0.66	0.58	11.3	0.67	0.50	11.3
ITEM 33	0.47	0.30	13.3	0.48	0.31	13.2	0.45	0.29	13.5
ITEM 34	0.51	0.59	12.9	0.50	0.59	13.0	0.53	0.60	12.7
ITEM 35	0.56	0.49	12.4	0.59	0.51	12.1	0.54	0.46	12.6
ITEM 36	0.40	0.63	14.0	0.42	0.64	13.8	0.38	0.61	14.2
ITEM 37	0.45	0.69	13.5	0.47	0.70	13.3	0.44	0.69	13.6
ITEM 38	0.42	0.31	13.8	0.44	0.33	13.6	0.40	0.28	14.0
ITEM 39	0.39	0.70	14.1	0.38	0.67	14.2	0.39	0.72	14.1
ITEM 40	0.20	0.61	15.3	0.20	0.58	15.3	0.19	0.64	15.3
COLUMN MEAN	0.54	0.58	12.5	0.55	0.58	12.5	0.54	0.57	12.6
COLUMN S.D.	0.13	0.11	1.3	0.13	0.11	1.3	0.13	0.12	1.4

SAMPLE SIZE	23647	11649	11801
POPULATION ESTIMATE	3000360	1491770	1489512
COEFFICIENT ALPHA	0.90	0.90	0.90
SPLIT HALF RELIABILITY	0.90	0.91	0.90

	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.
FORMULA SCORE	16.8	11.32	16.2	11.45	15.8	11.18
NUMBER RIGHT	21.6	8.72	21.7	8.84	21.5	8.60
NUMBER WRONG	17.5	8.38	17.3	8.46	17.4	8.31
NUMBER OMITS	0.8	2.18	0.8	2.19	0.7	2.17
NUMBER NOT REACHED	0.2	1.47	0.2	1.48	0.2	1.16

Source: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988: Base Year Survey .

Appendix A-2--(continued)

Item Analysis Statistics, Mathematics

	TOTAL			ASIAN			HISPANIC			BLACK			WHITE			AMERICAN INDIAN		
	P+	RBIS	DELTA	P+	RBIS	DELTA	P+	RBIS	DELTA	P+	RBIS	DELTA	P+	RBIS	DELTA	P+	RBIS	DELTA
ITEM 1	0.71	0.69	10.0	0.77	0.76	10.0	0.68	0.64	12.0	0.54	0.61	12.6	0.75	0.68	10.3	0.56	0.59	12.4
ITEM 2	0.50	0.60	13.0	0.56	0.65	12.4	0.38	0.53	14.2	0.31	0.51	14.9	0.55	0.58	12.5	0.31	0.45	15.0
ITEM 3	0.47	0.27	13.3	0.51	0.37	12.9	0.42	0.21	13.8	0.44	0.23	13.6	0.49	0.27	13.1	0.42	0.07	13.8
ITEM 4	0.49	0.65	13.1	0.58	0.72	12.1	0.36	0.57	14.5	0.36	0.58	14.4	0.53	0.65	12.7	0.32	0.54	14.9
ITEM 5	0.51	0.65	12.9	0.61	0.79	11.9	0.41	0.57	14.0	0.35	0.56	14.5	0.55	0.64	12.5	0.37	0.55	14.4
ITEM 6	0.48	0.65	13.5	0.49	0.52	13.1	0.36	0.33	14.5	0.37	0.35	14.4	0.48	0.48	13.2	0.36	0.26	14.5
ITEM 7	0.41	0.69	13.9	0.58	0.78	13.0	0.28	0.63	15.3	0.27	0.63	15.4	0.46	0.68	13.5	0.24	0.42	15.8
ITEM 8	0.37	0.59	14.4	0.53	0.67	12.7	0.28	0.44	15.4	0.23	0.47	15.9	0.48	0.59	14.0	0.25	0.37	15.7
ITEM 9	0.44	0.51	13.6	0.55	0.63	12.5	0.34	0.34	14.4	0.35	0.32	14.6	0.46	0.55	13.4	0.34	0.17	14.5
ITEM 10	0.41	0.48	13.9	0.42	0.62	13.8	0.28	0.46	15.3	0.26	0.49	15.6	0.46	0.59	13.4	0.27	0.43	15.5
ITEM 11	0.35	0.54	14.5	0.43	0.63	13.7	0.28	0.42	15.3	0.22	0.35	16.1	0.39	0.55	14.1	0.28	0.34	15.4
ITEM 12	0.44	0.66	13.6	0.55	0.72	12.5	0.35	0.63	14.5	0.32	0.60	14.9	0.47	0.66	13.3	0.31	0.69	15.0
ITEM 13	0.52	0.70	12.8	0.63	0.75	11.7	0.41	0.63	13.9	0.38	0.60	14.3	0.57	0.70	12.3	0.36	0.55	14.5
ITEM 14	0.50	0.69	13.0	0.61	0.72	11.9	0.38	0.65	14.2	0.37	0.60	14.3	0.54	0.69	12.6	0.34	0.63	14.7
ITEM 15	0.71	0.58	10.8	0.76	0.69	10.2	0.66	0.44	11.4	0.64	0.44	11.5	0.73	0.52	10.6	0.65	0.41	11.4
ITEM 16	0.79	0.49	9.8	0.82	0.56	9.3	0.75	0.49	10.3	0.73	0.42	10.5	0.81	0.49	9.6	0.71	0.55	10.8
ITEM 17	0.70	0.46	10.9	0.73	0.52	10.5	0.61	0.40	11.8	0.61	0.36	11.9	0.73	0.46	10.6	0.58	0.31	12.2
ITEM 18	0.52	0.64	12.8	0.63	0.75	11.7	0.42	0.58	13.8	0.38	0.54	14.2	0.55	0.65	12.5	0.39	0.47	14.1
ITEM 19	0.79	0.58	9.7	0.84	0.65	8.6	0.72	0.53	10.7	0.73	0.58	10.5	0.81	0.59	9.4	0.74	0.48	10.5
ITEM 20	0.79	0.58	9.6	0.85	0.57	8.8	0.75	0.53	10.3	0.72	0.44	10.7	0.81	0.49	9.8	0.69	0.36	11.0
ITEM 21	0.69	0.55	11.0	0.69	0.57	11.0	0.57	0.34	12.3	0.53	0.44	12.7	0.74	0.54	10.4	0.56	0.54	12.4
ITEM 22	0.68	0.71	11.1	0.76	0.78	10.1	0.55	0.66	12.5	0.46	0.64	13.5	0.75	0.70	10.4	0.52	0.69	12.8
ITEM 23	0.65	0.48	11.5	0.68	0.38	11.2	0.54	0.46	12.4	0.52	0.49	12.8	0.69	0.41	11.1	0.55	0.53	12.5
ITEM 24	0.59	0.57	12.1	0.63	0.64	11.7	0.50	0.50	13.0	0.43	0.51	13.7	0.63	0.54	11.6	0.45	0.57	13.6
ITEM 25	0.65	0.65	11.4	0.78	0.65	10.9	0.53	0.60	12.7	0.36	0.57	14.6	0.73	0.61	10.6	0.46	0.62	13.4
ITEM 26	0.62	0.59	11.7	0.65	0.72	11.4	0.55	0.48	12.5	0.49	0.48	13.1	0.66	0.61	11.3	0.48	0.58	13.2
ITEM 27	0.68	0.76	12.0	0.67	0.81	11.3	0.47	0.67	13.3	0.38	0.67	14.2	0.66	0.76	11.3	0.37	0.69	14.3
ITEM 28	0.56	0.59	12.4	0.61	0.59	11.9	0.44	0.54	13.6	0.37	0.48	14.3	0.62	0.58	11.8	0.44	0.58	13.6
ITEM 29	0.52	0.66	12.8	0.68	0.75	12.0	0.44	0.59	13.6	0.37	0.59	14.4	0.56	0.65	12.4	0.38	0.57	14.2
ITEM 30	0.52	0.54	12.8	0.56	0.61	12.2	0.44	0.49	13.6	0.43	0.49	13.7	0.55	0.54	12.5	0.42	0.46	13.8
ITEM 31	0.59	0.67	12.1	0.69	0.72	11.0	0.47	0.61	13.3	0.37	0.56	14.3	0.65	0.66	11.5	0.45	0.64	13.5
ITEM 32	0.66	0.84	11.3	0.71	0.66	10.8	0.60	0.47	12.8	0.53	0.39	12.7	0.70	0.55	10.9	0.54	0.46	12.6
ITEM 33	0.47	0.30	13.3	0.51	0.48	12.9	0.42	0.32	13.8	0.37	0.26	14.3	0.49	0.27	13.1	0.43	0.25	13.7
ITEM 34	0.51	0.59	12.9	0.65	0.69	11.4	0.43	0.46	13.7	0.39	0.52	14.2	0.55	0.60	12.5	0.41	0.51	13.9
ITEM 35	0.56	0.49	12.4	0.61	0.49	11.9	0.47	0.39	13.3	0.44	0.32	13.6	0.60	0.49	12.0	0.42	0.47	13.8
ITEM 36	0.48	0.63	14.0	0.51	0.70	12.9	0.30	0.57	15.0	0.24	0.45	15.8	0.44	0.62	13.6	0.27	0.37	15.5
ITEM 37	0.48	0.69	13.8	0.47	0.69	13.3	0.29	0.70	15.2	0.21	0.64	16.2	0.52	0.66	12.8	0.28	0.61	15.4
ITEM 38	0.42	0.31	13.8	0.46	0.48	13.4	0.39	0.27	14.1	0.37	0.21	14.3	0.44	0.31	13.6	0.36	0.29	14.4
ITEM 39	0.39	0.79	14.1	0.51	0.79	12.9	0.25	0.63	15.6	0.23	0.55	15.9	0.43	0.69	13.7	0.24	0.64	15.9
ITEM 40	0.28	0.61	15.3	0.43	0.69	13.5	0.21	0.52	16.2	0.17	0.52	16.8	0.31	0.68	15.0	0.21	0.49	16.3
COLUMN MEAN	0.54	0.58	12.5	0.61	0.64	11.8	0.46	0.51	13.5	0.41	0.49	14.0	0.58	0.57	12.1	0.42	0.48	13.9
COLUMN S.D.	0.13	0.11	1.3	0.11	0.11	1.3	0.13	0.11	1.4	0.14	0.11	1.5	0.13	0.11	1.4	0.13	0.14	1.4

	TOTAL	ASIAN	HISPANIC	BLACK	WHITE	AMERICAN INDIAN
SAMPLE SIZE	23647	1495	2995	2864	15760	307
POPULATION ESTIMATE	3809368	105333	303593	390442	2127458	43183
COEFFICIENT ALPHA	0.90	0.92	0.86	0.84	0.89	0.84
SPLIT HALF RELIABILITY	0.90	0.93	0.86	0.85	0.90	0.85

	TOTAL		ASIAN		HISPANIC		BLACK		WHITE		AMERICAN INDIAN	
	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.
FORMULA SCORE	16.0	11.32	19.7	12.23	11.1	9.88	8.9	9.06	16.0	11.05	9.5	9.13
NUMBER RIGHT	21.6	8.72	24.4	9.45	17.8	7.71	16.1	7.66	23.2	8.48	16.6	7.81
NUMBER WRONG	17.5	8.38	14.7	9.05	20.9	7.47	22.2	7.00	16.1	8.22	22.1	7.10
NUMBER OMITS	0.8	2.18	0.7	2.17	1.0	2.82	1.2	2.88	0.6	1.85	1.0	2.56
NUMBER NOT REACHED	0.2	1.47	0.2	1.68	0.4	2.17	0.5	2.34	0.1	1.05	0.3	1.43

Source: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988: Base Year Survey .

Appendix A-2--(continued)

Item Analysis Statistics, Mathematics

	HISPANIC MALE			HISPANIC FEMALE			BLACK MALE			BLACK FEMALE			WHITE MALE			WHITE FEMALE		
	P+	RBIS	DELTA	P+	RBIS	DELTA	P+	RBIS	DELTA	P+	RBIS	DELTA	P+	RBIS	DELTA	P+	RBIS	DELTA
ITEM 1	0.60	0.65	12.8	0.59	0.63	12.1	0.50	0.62	13.0	0.57	0.61	12.3	0.74	0.68	10.5	0.77	0.68	10.0
ITEM 2	0.40	0.52	14.8	0.35	0.55	14.5	0.32	0.53	14.8	0.30	0.56	15.1	0.55	0.58	12.5	0.54	0.57	12.6
ITEM 3	0.41	0.21	13.9	0.43	0.22	13.7	0.41	0.25	13.9	0.47	0.28	13 3	0.48	0.28	13.2	0.50	0.27	13.0
ITEM 4	0.36	0.44	14.4	0.36	0.50	14.5	0.34	0.60	14.6	0.30	0.64	14.2	0.53	0.67	12.7	0.53	0.62	12.7
ITEM 5	0.43	0.56	13.7	0.38	0.59	14.2	0.35	0.51	14.5	0.36	0.62	14.5	0.56	0.65	12.4	0.54	0.64	12.6
ITEM 6	0.35	0.31	14.5	0.36	0.36	14.4	0.36	0.35	14.4	0.37	0.35	14.4	0.47	0.44	13.3	0.48	0.47	13.2
ITEM 7	0.28	0.64	15.3	0.28	0.59	15.3	0.27	0.61	15.4	0.28	0.65	15.4	0.45	0.67	13.5	0.44	0.69	13.4
ITEM 8	0.28	0.39	15.3	0.27	0.48	15.4	0.24	0.49	15.9	0.23	0.46	16.0	0.39	0.56	14.1	0.41	0.62	13.9
ITEM 9	0.48	0.38	14.0	0.33	0.32	14.8	0.38	0.32	14.2	0.32	0.32	14.9	0.48	0.54	13.2	0.45	0.55	13.6
ITEM 10	0.26	0.49	15.6	0.30	0.44	15.1	0.25	0.50	15.7	0.26	0.49	15.6	0.44	0.63	13.4	0.44	0.55	13.4
ITEM 11	0.31	0.39	14.9	0.25	0.44	15.0	0.24	0.34	15.8	0.21	0.35	16.3	0.41	0.53	14.0	0.37	0.57	14.3
ITEM 12	0.34	0.64	14.6	0.36	0.63	14.4	0.29	0.61	15.2	0.35	0.60	14.6	0.45	0.65	13.5	0.50	0.67	13.0
ITEM 13	0.43	0.64	13.7	0.38	0.62	14.2	0.38	0.60	14.2	0.38	0.60	14.3	0.58	0.70	12.2	0.55	0.70	12.5
ITEM 14	0.42	0.65	13.8	0.34	0.65	14.6	0.38	0.60	14.2	0.36	0.61	14.5	0.56	0.68	12.4	0.52	0.70	12.8
ITEM 15	0.66	0.47	11.4	0.66	0.42	11.4	0.62	0.48	11.7	0.66	0.40	11.3	0.71	0.54	10.8	0.74	0.50	10.4
ITEM 16	0.76	0.50	10.2	0.74	0.48	10.4	0.71	0.46	10.8	0.75	0.39	10.3	0.81	0.56	9.5	0.80	0.43	9.6
ITEM 17	0.62	0.43	11.7	0.68	0.36	11.9	0.58	0.39	12.1	0.64	0.33	11.6	0.73	0.49	10.5	0.73	0.42	10.6
ITEM 18	0.45	0.52	13.5	0.48	0.48	14.0	0.37	0.56	14.4	0.39	0.52	14.1	0.54	0.64	12.6	0.56	0.67	12.4
ITEM 19	0.78	0.50	10.9	0.73	0.48	10.6	0.78	0.50	11.0	0.77	0.50	10.1	0.88	0 61	9.6	0.83	0.58	9.2
ITEM 20	0.72	0.61	10.6	0.78	0.45	9.9	0.67	0.49	11.2	0.76	0.43	10.2	0.78	0.[illegible]	9.9	0.84	0.43	9.0
ITEM 21	0.62	0.58	11.7	0.51	0.49	12.9	0.57	0.45	12.3	0.49	0.42	13.1	0.78	0.58	9.9	0.70	0.51	10.9
ITEM 22	0.58	0.68	12.2	0.51	0.63	12.9	0.46	0.66	13.4	0.45	0.63	13.5	0.76	0.70	10.2	0.73	0.70	10.6
ITEM 23	0.58	0.44	12.2	0.54	0.47	12.6	0.54	0.46	12.6	0.58	0.52	13.0	0.68	0.41	11.1	0.69	0.41	11.0
ITEM 24	0.49	0.64	13.1	0.51	0.52	12.9	0.44	0.51	13.6	0.42	0.52	13.8	0.63	0.54	11.6	0.64	0.53	11.6
ITEM 25	0.68	0.58	12.0	0.46	0.62	13.4	0.41	0.56	13.9	0.28	0.61	15.3	0.78	0.58	9.9	0.68	0.65	11.2
ITEM 26	0.58	0.46	12.2	0.53	0.48	12.7	0.49	0.45	13.1	0.49	0.50	13.1	0.67	0.61	11.2	0.65	0.60	11.5
ITEM 27	0.51	0.66	12.9	0.44	0.66	13.6	0.41	0.66	13.9	0.36	0.68	14.5	0.67	0.76	11.2	0.65	0.76	11.4
ITEM 28	0.44	0.59	13.6	0.43	0.58	13.7	0.34	0.67	14.7	0.40	0.49	14.0	0.59	0.62	12.1	0.65	0.54	11.5
ITEM 29	0.43	0.62	13.7	0.44	0.56	13.6	0.34	0.68	14.6	0.39	0.58	14.1	0.56	0.67	12.6	0.57	0.64	12.3
ITEM 30	0.44	0.54	13.6	0.44	0.44	13.6	0.39	0.47	14.1	0.44	0.51	13.4	0.53	0.56	12.7	0.57	0.51	12.3
ITEM 31	0.49	0.60	13.1	0.44	0.61	13.6	0.38	0.55	14.2	0.36	0.57	14.4	0.67	0.66	11.3	0.63	0.66	11.6
ITEM 32	0.59	0.53	12.0	0.60	0.41	12.0	0.52	0.37	12.8	0.53	0.40	12.7	0.70	0.61	11.0	0.70	0.49	10.9
ITEM 33	0.43	0.33	13.7	0.40	0.30	14.0	0.37	0.21	14.3	0.37	0.29	14.3	0.51	0.28	12.9	0.47	0.26	13.3
ITEM 34	0.43	0.42	13.7	0.42	0.51	13.8	0.38	0.50	14.2	0.39	0.53	14.1	0.53	0.61	12.7	0.57	0.59	12.3
ITEM 35	0.58	0.43	13.0	0.43	0.34	13.7	0.46	0.34	13.4	0.43	0.31	13.8	0.62	0.53	11.7	0.58	0.46	12.2
ITEM 36	0.32	0.63	14.8	0.29	0.52	15.2	0.25	0.42	15.7	0.24	0.48	15.9	0.47	0.64	13.3	0.42	0.61	13.8
ITEM 37	0.32	0.74	14.9	0.27	0.65	15.5	0.21	0.61	16.2	0.20	0.67	16.3	0.54	0.67	12.6	0.51	0.65	12.9
ITEM 38	0.41	0.27	13.9	0.37	0.26	14.4	0.38	0.21	14.2	0.36	0.22	14.4	0.46	0.34	13.4	0.41	0.29	13.9
ITEM 39	0.27	0.62	15.4	0.24	0.62	15.9	0.23	0.49	16.0	0.23	0.61	15.9	0.42	0.66	13.8	0.44	0.72	13.6
ITEM 40	0.23	0.53	15.9	0.19	0.58	16.5	0.17	0.49	16.8	0.17	0.55	16.8	0.30	0.57	15.1	0.32	0.64	14.9
COLUMN MEAN	0.46	0.53	13.4	0.43	0.49	13.7	0.40	0.48	14.0	0.41	0.49	14.0	0.58	0.58	12.1	0.58	0.56	12.1
COLUMN S.D.	0.13	0.12	1.4	0.14	0.11	1.5	0.13	0.11	1.4	0.15	0.12	1.6	0.13	0.11	1.4	0.13	0.12	1.4

SAMPLE SIZE	1429			1545			1383			1442			7825			7821		
POPULATION ESTIMATE	150434			151157			191201			196786			1059771			1054962		
COEFFICIENT ALPHA	0.87			0.85			0.84			0.84			0.90			0.89		
SPLIT HALF RELIABILITY	0.89			0.86			0.86			0.85			0.90			0.90		
	MEAN	S.D.		MEAN	S.D.		MEAN	S.D.		MEAN	S.D.		MEAN	S.D.		MEAN	S.D.	
FORMULA SCORE	11.7	10.23		10.4	9.42		8.6	9.05		9.0	9.08		18.1	11.19		17.9	10.90	
NUMBER RIGHT	18.3	8.01		17.3	7.32		16.0	7.10		16.2	7.04		23.3	8.58		23.1	8.35	
NUMBER WRONG	20.3	7.69		21.4	7.18		22.2	6.90		22.2	7.01		16.0	8.30		16.2	8.12	
NUMBER OMITS	1.0	2.68		1.1	2.96		1.3	3.02		1.1	2.59		0.6	1.79		0.6	1.89	
NUMBER NOT REACHED	0.5	2.68		0.3	1.53		0.6	2.59		0.5	2.03		0.1	1.19		0.1	0.88	

Source: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988: Base Year Survey .

Appendix A-3

Item Analysis Statistics, Science

	TOTAL			MALE			FEMALE		
	P+	RBIS	DELTA	P+	RBIS	DELTA	P+	RBIS	DELTA
ITEM 1	0.70	0.57	10.9	0.69	0.60	11.0	0.70	0.55	10.9
ITEM 2	0.79	0.51	9.8	0.80	0.60	9.6	0.77	0.41	10.1
ITEM 3	0.64	0.48	11.6	0.65	0.49	11.5	0.63	0.48	11.6
ITEM 4	0.67	0.45	11.3	0.63	0.47	11.6	0.70	0.45	10.9
ITEM 5	0.76	0.71	10.2	0.77	0.78	10.0	0.74	0.64	10.4
ITEM 6	0.76	0.67	10.2	0.76	0.71	10.2	0.76	0.62	10.2
ITEM 7	0.65	0.50	11.4	0.70	0.58	10.9	0.61	0.42	11.9
ITEM 8	0.57	0.46	12.3	0.61	0.50	11.9	0.54	0.42	12.6
ITEM 9	0.64	0.51	11.6	0.64	0.52	11.6	0.64	0.51	11.5
ITEM 10	0.53	0.53	12.7	0.54	0.55	12.6	0.53	0.51	12.7
ITEM 11	0.48	0.42	13.2	0.50	0.46	13.0	0.46	0.36	13.4
ITEM 12	0.66	0.56	11.3	0.70	0.59	10.9	0.62	0.54	11.7
ITEM 13	0.72	0.54	10.6	0.70	0.59	10.9	0.75	0.50	10.3
ITEM 14	0.53	0.65	12.7	0.58	0.66	12.2	0.49	0.64	13.1
ITEM 15	0.39	0.47	14.1	0.37	0.47	14.3	0.41	0.49	13.9
ITEM 16	0.46	0.42	13.4	0.46	0.43	13.4	0.46	0.41	13.4
ITEM 17	0.42	0.49	13.8	0.45	0.53	13.5	0.39	0.45	14.1
ITEM 18	0.45	0.54	13.5	0.49	0.56	13.1	0.41	0.52	13.9
ITEM 19	0.42	0.51	13.8	0.43	0.52	13.7	0.41	0.49	13.9
ITEM 20	0.41	0.35	13.9	0.41	0.37	13.9	0.41	0.33	13.9
ITEM 21	0.42	0.39	13.8	0.44	0.42	13.6	0.40	0.35	14.0
ITEM 22	0.37	0.38	14.3	0.35	0.40	14.6	0.39	0.37	14.1
ITEM 23	0.39	0.27	14.1	0.40	0.30	14.0	0.39	0.24	14.1
ITEM 24	0.32	0.56	14.8	0.33	0.56	14.7	0.32	0.55	14.9
ITEM 25	0.22	0.37	16.1	0.21	0.35	16.2	0.23	0.39	16.0
COLUMN MEAN	0.53	0.49	12.6	0.54	0.52	12.5	0.53	0.47	12.7
COLUMN S.D.	0.15	0.10	1.6	0.16	0.11	1.7	0.15	0.10	1.6

SAMPLE SIZE	23623	11664	11783
POPULATION ESTIMATE	2993973	1489380	1485637
COEFFICIENT ALPHA	0.75	0.78	0.72
SPLIT HALF RELIABILITY	0.77	0.79	0.73

	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.
FORMULA SCORE	9.9	5.83	10.2	6.10	9.6	5.52
NUMBER RIGHT	13.3	4.52	13.6	4.74	13.1	4.29
NUMBER WRONG	1[illegible].2	4.48	11.0	4.67	11.5	4.26
NUMBER OMITS	0.3	0.96	0.3	0.97	0.3	0.95
NUMBER NOT REACHED	0.1	0.98	0.1	1.05	0.1	0.91

Source: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988: Base Year Survey .

Appendix A-3--(continued)

Item Analysis Statistics, Science

	TOTAL			ASIAN			HISPANIC			BLACK			WHITE			AMERICAN INDIAN		
	P+	RBIS	DELTA	P+	RBIS	DELTA	P+	RBIS	DELTA	P+	RBIS	DELTA	P+	RBIS	DELTA	P+	RBIS	DELTA
ITEM 1	0.70	0.57	10.9	0.68	0.59	11.1	0.63	0.48	11.6	0.51	0.45	12.9	0.75	0.57	10.4	0.55	0.50	12.5
ITEM 2	0.79	0.51	9.8	0.81	0.55	9.5	0.72	0.49	10.6	0.69	0.44	11.0	0.81	0.49	9.4	0.65	0.57	11.4
ITEM 3	0.64	0.48	11.6	0.68	0.52	11.2	0.57	0.46	12.3	0.53	0.40	12.7	0.67	0.47	11.2	0.52	0.43	12.8
ITEM 4	0.67	0.45	11.3	0.66	0.42	11.3	0.62	0.38	11.8	0.57	0.40	12.3	0.69	0.45	11.0	0.55	0.59	12.5
ITEM 5	0.76	0.71	10.2	0.78	0.70	10.0	0.67	0.64	11.2	0.58	0.62	12.2	0.80	0.71	9.6	0.62	0.69	11.8
ITEM 6	0.76	0.67	10.2	0.76	0.69	10.1	0.65	0.60	11.4	0.65	0.58	11.5	0.80	0.67	9.7	0.59	0.66	12.1
ITEM 7	0.65	0.50	11.4	0.70	0.46	10.9	0.61	0.48	11.9	0.55	0.46	12.5	0.68	0.49	11.1	0.54	0.55	12.6
ITEM 8	0.57	0.46	12.3	0.53	0.52	12.7	0.48	0.46	13.2	0.48	0.39	13.2	0.61	0.45	11.9	0.46	0.51	13.4
ITEM 9	0.64	0.51	11.6	0.66	0.54	11.3	0.56	0.48	12.4	0.53	0.46	12.7	0.68	0.50	11.2	0.49	0.49	13.1
ITEM 10	0.53	0.53	12.7	0.55	0.58	12.4	0.41	0.46	13.9	0.43	0.39	13.7	0.57	0.53	12.2	0.39	0.52	14.1
ITEM 11	0.48	0.42	13.2	0.53	0.39	12.7	0.42	0.44	13.8	0.40	0.36	14.0	0.50	0.41	13.0	0.35	0.39	14.6
ITEM 12	0.66	0.56	11.3	0.70	0.61	10.9	0.57	0.54	12.3	0.52	0.47	12.8	0.70	0.55	10.9	0.58	0.45	12.2
ITEM 13	0.72	0.54	10.6	0.77	0.50	10.1	0.66	0.52	11.3	0.61	0.50	11.9	0.75	0.53	10.3	0.60	0.62	12.0
ITEM 14	0.53	0.65	12.7	0.55	0.67	12.5	0.36	0.53	14.4	0.25	0.48	15.7	0.61	0.63	11.9	0.33	0.51	14.8
ITEM 15	0.39	0.47	14.1	0.45	0.47	13.5	0.37	0.45	14.3	0.28	0.43	15.4	0.41	0.46	13.9	0.27	0.49	15.4
ITEM 16	0.46	0.42	13.4	0.49	0.47	13.1	0.43	0.31	13.7	0.39	0.32	14.1	0.48	0.44	13.2	0.34	0.37	14.7
ITEM 17	0.42	0.49	13.8	0.45	0.54	13.5	0.34	0.39	14.7	0.32	0.30	14.9	0.45	0.51	13.5	0.32	0.35	14.8
ITEM 18	0.45	0.54	13.5	0.45	0.55	13.5	0.34	0.41	14.7	0.30	0.34	15.1	0.50	0.55	13.0	0.34	0.48	14.6
ITEM 19	0.42	0.51	13.8	0.49	0.53	13.1	0.33	0.39	14.7	0.31	0.45	15.0	0.46	0.50	13.5	0.28	0.47	15.4
ITEM 20	0.41	0.35	13.9	0.44	0.45	13.6	0.36	0.28	14.4	0.36	0.30	14.4	0.43	0.36	13.7	0.34	0.17	14.7
ITEM 21	0.42	0.39	13.8	0.47	0.41	13.3	0.36	0.29	14.4	0.36	0.27	14.4	0.44	0.41	13.6	0.38	0.21	14.3
ITEM 22	0.37	0.38	14.3	0.44	0.39	13.6	0.33	0.31	14.8	0.29	0.34	15.2	0.38	0.38	14.2	0.27	0.48	15.4
ITEM 23	0.39	0.27	14.1	0.43	0.35	13.7	0.35	0.20	14.5	0.34	0.25	14.7	0.41	0.27	13.9	0.41	0.29	13.9
ITEM 24	0.32	0.56	14.8	0.34	0.58	14.6	0.24	0.53	15.8	0.20	0.51	16.4	0.36	0.54	14.4	0.18	0.35	16.7
ITEM 25	0.22	0.37	16.1	0.24	0.35	15.9	0.18	0.33	16.7	0.16	0.32	17.0	0.24	0.36	15.9	0.14	0.39	17.2
COLUMN MEAN	0.53	0.49	12.6	0.56	0.51	12.3	0.46	0.43	13.4	0.42	0.41	13.8	0.57	0.49	12.3	0.42	0.46	13.9
COLUMN S.D.	0.15	0.10	1.6	0.15	0.10	1.6	0.15	0.10	1.6	0.14	0.09	1.5	0.16	0.10	1.7	0.14	0.13	1.5

SAMPLE SIZE	23623	1492	2989	2849	15760	307
POPULATION ESTIMATE	2993973	105061	302672	385339	2127441	43183
COEFFICIENT ALPHA	0.75	0.77	0.67	0.62	0.74	0.71
SPLIT HALF RELIABILITY	0.77	0.78	0.69	0.65	0.76	0.72

	TOTAL		ASIAN		HISPANIC		BLACK		WHITE		AMERICAN INDIAN	
	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.
FORMULA SCORE	9.9	5.83	10.8	6.05	7.5	5.19	6.3	4.81	10.9	5.68	6.2	5.43
NUMBER RIGHT	13.3	4.52	14.0	4.71	11.5	4.05	10.5	3.76	14.2	4.39	10.4	4.28
NUMBER WRONG	11.2	4.48	10.5	4.67	12.9	4.07	13.7	3.93	10.5	4.36	13.8	4.25
NUMBER OMITS	0.3	0.96	0.3	0.93	0.4	1.03	0.4	1.10	0.3	0.89	0.5	1.66
NUMBER NOT REACHED	0.1	0.98	0.2	1.25	0.2	1.29	0.3	1.70	0.1	0.66	0.3	1.53

Source: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988: Base Year Survey .

Appendix A-3--(continued)

Item Analysis Statistics, Science

	HISPANIC MALE			HISPANIC FEMALE			BLACK MALE			BLACK FEMALE			WHITE MALE			WHITE FEMALE		
	P+	RBIS	DELTA	P+	RBIS	DELTA	P+	RBIS	DELTA	P+	RBIS	DELTA	P+	RBIS	DELTA	P+	RBIS	DELTA
ITEM 1	0.62	0.52	11.7	0.64	0.45	11.5	0.50	0.45	13.0	0.52	0.47	12.8	0.74	0.60	10.4	0.75	0.54	10.3
ITEM 2	0.73	0.54	10.5	0.71	0.44	10.8	0.69	0.49	11.0	0.70	0.38	10.9	0.84	0.60	9.1	0.79	0.39	9.8
ITEM 3	0.59	0.46	12.1	0.56	0.47	12.4	0.54	0.41	12.6	0.52	0.38	12.8	0.67	0.49	11.2	0.67	0.46	11.2
ITEM 4	0.60	0.37	12.0	0.65	0.41	11.5	0.54	0.42	12.6	0.59	0.39	12.1	0.66	0.48	11.3	0.73	0.44	10.6
ITEM 5	0.69	0.71	11.0	0.65	0.56	11.4	0.60	0.67	12.0	0.56	0.58	12.4	0.82	0.79	9.4	0.79	0.63	9.8
ITEM 6	0.68	0.63	11.1	0.62	0.57	11.8	0.64	0.61	11.6	0.65	0.56	11.4	0.80	0.73	9.7	0.80	0.61	9.6
ITEM 7	0.66	0.53	11.3	0.56	0.43	12.4	0.57	0.54	12.3	0.53	0.37	12.7	0.73	0.58	10.6	0.63	0.41	11.7
ITEM 8	0.52	0.48	12.8	0.44	0.42	13.6	0.51	0.44	12.9	0.46	0.33	13.4	0.64	0.49	11.5	0.57	0.41	12.3
ITEM 9	0.55	0.50	12.5	0.57	0.46	12.3	0.54	0.49	12.6	0.52	0.43	12.8	0.67	0.50	11.2	0.68	0.50	11.1
ITEM 10	0.43	0.50	13.7	0.39	0.41	14.1	0.43	0.37	13.7	0.42	0.41	13.8	0.58	0.56	12.2	0.57	0.50	12.3
ITEM 11	0.43	0.48	13.7	0.40	0.39	14.0	0.39	0.43	14.1	0.42	0.29	13.8	0.53	0.45	12.7	0.47	0.36	13.3
ITEM 12	0.61	0.56	11.9	0.52	0.52	12.8	0.56	0.48	12.4	0.48	0.46	13.2	0.74	0.58	10.4	0.67	0.52	11.3
ITEM 13	0.62	0.56	11.7	0.69	0.50	11.0	0.56	0.50	12.4	0.66	0.49	11.3	0.73	0.58	10.5	0.77	0.49	10.0
ITEM 14	0.40	0.58	14.0	0.32	0.45	14.8	0.29	0.45	15.2	0.21	0.51	16.2	0.66	0.64	11.3	0.56	0.63	12.4
ITEM 15	0.37	0.44	14.4	0.37	0.46	14.3	0.27	0.43	15.4	0.28	0.44	15.3	0.39	0.47	14.1	0.44	0.47	13.6
ITEM 16	0.41	0.35	13.9	0.45	0.29	13.6	0.39	0.31	14.2	0.40	0.33	14.0	0.49	0.45	13.1	0.48	0.43	13.2
ITEM 17	0.35	0.41	14.6	0.33	0.36	14.8	0.32	0.34	14.9	0.31	0.25	15.0	0.48	0.54	13.2	0.42	0.48	13.8
ITEM 18	0.35	0.44	14.5	0.32	0.37	14.8	0.31	0.27	15.0	0.28	0.42	15.3	0.54	0.58	12.6	0.45	0.52	13.5
ITEM 19	0.34	0.38	14.6	0.32	0.40	14.9	0.31	0.45	15.0	0.30	0.45	15.0	0.46	0.52	13.4	0.45	0.48	13.6
ITEM 20	0.37	0.32	14.4	0.36	0.23	14.4	0.36	0.28	14.5	0.37	0.31	14.4	0.43	0.39	13.7	0.42	0.34	13.8
ITEM 21	0.38	0.30	14.2	0.34	0.28	14.6	0.36	0.32	14.5	0.37	0.22	14.4	0.46	0.44	13.4	0.42	0.38	13.8
ITEM 22	0.31	0.28	15.0	0.35	0.36	14.5	0.27	0.34	15.5	0.32	0.34	14.9	0.37	0.41	14.4	0.41	0.37	14.0
ITEM 23	0.36	0.24	14.4	0.34	0.15	14.6	0.34	0.25	14.7	0.34	0.23	14.7	0.41	0.30	13.9	0.40	0.24	14.0
ITEM 24	0.26	0.56	15.6	0.22	0.49	16.1	0.19	0.56	16.5	0.20	0.45	16.3	0.37	0.53	14.3	0.36	0.54	14.5
ITEM 25	0.17	0.36	16.9	0.20	0.33	16.4	0.14	0.30	17.3	0.17	0.34	16.8	0.23	0.34	16.0	0.25	0.38	15.8
COLUMN MEAN	0.47	0.46	13.3	0.45	0.41	13.5	0.42	0.42	13.8	0.42	0.39	13.8	0.58	0.52	12.1	0.56	0.46	12.4
COLUMN S.D.	0.15	0.11	1.6	0.15	0.10	1.6	0.14	0.10	1.6	0.14	0.09	1.6	0.16	0.11	1.8	0.16	0.09	1.7
SAMPLE SIZE	1431			1537			1375			1455			7827			7820		
POPULATION ESTIMATE	150344			150327			188257			194547			1060421			1054444		
COEFFICIENT ALPHA	0.71			0.62			0.65			0.58			0.77			0.70		
SPLIT HALF RELIABILITY	0.73			0.64			0.68			0.62			0.79			0.72		
	MEAN	S.D.		MEAN	S.D.		MEAN	S.D.		MEAN	S.D.		MEAN	S.D.		MEAN	S.D.	
FORMULA SCORE	7.8	5.48		7.2	4.86		6.3	4.99		6.3	4.61		11.3	5.94		10.6	5.38	
NUMBER RIGHT	11.8	4.26		11.3	3.80		10.5	3.91		10.5	3.60		14.4	4.59		13.9	4.16	
NUMBER WRONG	12.7	4.24		13.1	3.87		13.7	4.08		13.8	3.76		10.2	4.55		10.8	4.14	
NUMBER OMITS	0.4	0.97		0.4	1.08		0.5	1.12		0.4	1.09		0.3	0.92		0.3	0.86	
NUMBER NOT REACHED	0.2	1.21		0.2	1.33		0.4	1.89		0.3	1.49		0.1	0.68		0.1	0.64	

Source: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988: Base Year Survey .

Appendix A-4

Item Analysis Statistics, History/Citizenship/Geography

	TOTAL			MALE			FEMALE		
	P+	RBIS	DELTA	P+	RBIS	DELTA	P+	RBIS	DELTA
ITEM 1	0.80	0.58	9.7	0.79	0.58	9.7	0.80	0.58	9.6
ITEM 2	0.77	0.66	10.0	0.77	0.69	10.1	0.78	0.62	9.9
ITEM 3	0.90	0.76	7.9	0.88	0.79	8.2	0.91	0.73	7.6
ITEM 4	0.68	0.63	11.1	0.70	0.67	10.9	0.67	0.59	11.3
ITEM 5	0.86	0.66	8.7	0.87	0.64	8.5	0.85	0.68	8.8
ITEM 6	0.84	0.54	9.1	0.83	0.55	9.2	0.84	0.53	8.9
ITEM 7	0.91	0.85	7.7	0.90	0.86	7.8	0.91	0.85	7.6
ITEM 8	0.88	0.73	8.3	0.88	0.73	8.2	0.88	0.72	8.3
ITEM 9	0.91	0.85	7.6	0.91	0.85	7.6	0.91	0.86	7.5
ITEM 10	0.70	0.47	11.0	0.70	0.51	10.9	0.70	0.44	10.9
ITEM 11	0.59	0.63	12.1	0.63	0.66	11.7	0.55	0.59	12.5
ITEM 12	0.55	0.52	12.5	0.52	0.54	12.8	0.58	0.51	12.2
ITEM 13	0.58	0.58	12.2	0.61	0.63	11.9	0.55	0.53	12.5
ITEM 14	0.42	0.41	13.8	0.44	0.43	13.6	0.40	0.40	14.0
ITEM 15	0.47	0.59	13.3	0.48	0.62	13.2	0.46	0.55	13.4
ITEM 16	0.45	0.45	13.5	0.46	0.50	13.4	0.44	0.40	13.7
ITEM 17	0.83	0.64	9.1	0.84	0.68	9.0	0.83	0.60	9.2
ITEM 18	0.78	0.59	9.9	0.78	0.61	9.9	0.78	0.56	9.9
ITEM 19	0.76	0.73	10.1	0.74	0.77	10.4	0.79	0.69	9.8
ITEM 20	0.66	0.60	11.4	0.66	0.62	11.3	0.65	0.58	11.5
ITEM 21	0.66	0.59	11.4	0.73	0.66	10.5	0.59	0.54	12.1
ITEM 22	0.48	0.56	13.2	0.48	0.58	13.2	0.48	0.53	13.2
ITEM 23	0.48	0.48	13.2	0.48	0.52	13.2	0.47	0.45	13.3
ITEM 24	0.54	0.54	12.6	0.54	0.58	12.6	0.54	0.49	12.6
ITEM 25	0.47	0.46	13.3	0.46	0.45	13.4	0.48	0.48	13.2
ITEM 26	0.49	0.52	13.1	0.51	0.54	12.9	0.46	0.49	13.4
ITEM 27	0.51	0.60	12.9	0.52	0.63	12.8	0.51	0.58	12.9
ITEM 28	0.43	0.46	13.7	0.47	0.49	13.3	0.39	0.43	14.1
ITEM 29	0.35	0.35	14.5	0.35	0.32	14.5	0.35	0.38	14.5
ITEM 30	0.25	0.28	15.8	0.26	0.26	15.5	0.23	0.29	16.0
COLUMN MEAN	0.63	0.58	11.4	0.64	0.60	11.3	0.63	0.56	11.5
COLUMN S.D.	0.18	0.13	2.2	0.18	0.13	2.1	0.19	0.13	2.3

SAMPLE SIZE	23536	11608	11753
POPULATION ESTIMATE	2984583	1484333	1481344
COEFFICIENT ALPHA	0.83	0.85	0.82
SPLIT HALF RELIABILITY	0.84	0.86	0.82

	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.
FORMULA SCORE	15.1	7.64	15.4	7.91	14.8	7.33
NUMBER RIGHT	18.9	5.53	19.2	5.75	18.7	5.29
NUMBER WRONG	10.8	5.41	10.5	5.60	11.0	5.20
NUMBER OMITS	0.2	0.92	0.2	0.87	0.2	0.95
NUMBER NOT REACHED	0.1	0.89	0.1	0.89	0.1	0.91

Source: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988: Base Year Survey .

Appendix A-4--(continued)

Item Analysis Statistics, History/Citizenship/Geography

	TOTAL			ASIAN			HISPANIC			BLACK			WHITE			AMERICAN INDIAN		
	P+	RBIS	DELTA	P+	RBIS	DELTA	P+	RBIS	DELTA	P+	RBIS	DELTA	P+	RBIS	DELTA	P+	RBIS	DELTA
ITEM 1	0.80	0.58	9.7	0.84	0.57	9.0	0.74	0.54	10.4	0.66	0.47	11.3	0.83	0.58	9.2	0.69	0.45	11.1
ITEM 2	0.77	0.66	10.0	0.75	0.72	10.3	0.64	0.60	11.6	0.73	0.58	10.5	0.81	0.67	9.5	0.65	0.58	11.5
ITEM 3	0.90	0.76	7.9	0.90	0.80	7.8	0.84	0.66	9.0	0.82	0.66	9.4	0.92	0.79	7.3	0.82	0.73	9.4
ITEM 4	0.68	0.63	11.1	0.63	0.62	11.6	0.50	0.54	13.0	0.54	0.57	12.6	0.74	0.62	10.4	0.55	0.59	12.5
ITEM 5	0.86	0.66	8.7	0.86	0.72	8.6	0.80	0.62	9.7	0.79	0.56	9.8	0.89	0.67	8.2	0.75	0.54	10.3
ITEM 6	0.84	0.54	9.1	0.85	0.64	8.9	0.75	0.54	10.3	0.78	0.53	9.9	0.86	0.50	8.7	0.79	0.62	9.7
ITEM 7	0.91	0.85	7.7	0.89	0.95	8.0	0.82	0.82	9.3	0.83	0.78	9.2	0.94	0.86	6.8	0.79	0.79	9.8
ITEM 8	0.88	0.73	8.3	0.87	0.80	8.4	0.79	0.70	9.7	0.83	0.67	9.2	0.91	0.72	7.7	0.79	0.67	9.7
ITEM 9	0.91	0.85	7.6	0.89	0.93	8.1	0.81	0.81	9.5	0.84	0.77	9.0	0.94	0.87	6.6	0.78	0.87	9.9
ITEM 10	0.70	0.47	11.0	0.70	0.58	10.9	0.67	0.41	11.2	0.62	0.38	11.8	0.72	0.49	10.7	0.62	0.42	11.8
ITEM 11	0.59	0.63	12.1	0.62	0.66	11.7	0.48	0.53	13.2	0.45	0.44	13.5	0.63	0.65	11.7	0.44	0.49	13.6
ITEM 12	0.55	0.52	12.5	0.64	0.52	11.5	0.47	0.48	13.3	0.46	0.46	13.5	0.57	0.53	12.3	0.44	0.34	13.6
ITEM 13	0.58	0.58	12.2	0.59	0.63	12.1	0.52	0.51	12.8	0.50	0.47	13.0	0.60	0.60	11.9	0.47	0.52	13.3
ITEM 14	0.42	0.41	13.8	0.56	0.52	12.3	0.49	0.43	13.1	0.35	0.34	14.6	0.41	0.43	13.9	0.33	0.32	14.8
ITEM 15	0.47	0.59	13.3	0.53	0.59	12.7	0.40	0.50	14.1	0.33	0.49	14.7	0.50	0.59	13.0	0.36	0.41	14.5
ITEM 16	0.45	0.45	13.5	0.54	0.48	12.6	0.38	0.42	14.2	0.36	0.31	14.5	0.47	0.47	13.3	0.38	0.21	14.2
ITEM 17	0.83	0.64	9.1	0.81	0.69	9.5	0.73	0.65	10.5	0.83	0.61	9.2	0.86	0.64	8.7	0.69	0.61	11.1
ITEM 18	0.78	0.59	9.9	0.80	0.61	9.6	0.70	0.53	10.9	0.68	0.50	11.1	0.81	0.59	9.5	0.62	0.65	11.8
ITEM 19	0.76	0.73	10.1	0.82	0.76	9.3	0.70	0.63	10.9	0.63	0.65	11.6	0.80	0.76	9.7	0.62	0.68	11.8
ITEM 20	0.66	0.60	11.4	0.65	0.65	11.4	0.54	0.53	12.6	0.52	0.48	12.8	0.70	0.60	10.8	0.50	0.61	13.0
ITEM 21	0.66	0.59	11.4	0.76	0.65	10.1	0.57	0.51	12.3	0.48	0.47	13.2	0.70	0.59	10.9	0.54	0.48	12.6
ITEM 22	0.48	0.56	13.2	0.57	0.54	12.3	0.44	0.44	13.6	0.34	0.42	14.7	0.51	0.58	12.9	0.38	0.37	14.3
ITEM 23	0.48	0.48	13.2	0.52	0.50	12.8	0.44	0.45	13.6	0.40	0.39	14.0	0.50	0.50	13.0	0.36	0.41	14.4
ITEM 24	0.54	0.54	12.6	0.56	0.52	12.4	0.47	0.47	13.3	0.45	0.49	13.5	0.57	0.54	12.3	0.42	0.47	13.9
ITEM 25	0.47	0.46	13.3	0.52	0.45	12.8	0.40	0.39	14.0	0.40	0.41	14.1	0.49	0.47	13.1	0.37	0.32	14.3
ITEM 26	0.49	0.52	13.1	0.50	0.46	13.0	0.37	0.41	14.3	0.32	0.31	14.8	0.54	0.53	12.6	0.35	0.41	14.6
ITEM 27	0.51	0.60	12.9	0.58	0.63	12.2	0.41	0.53	13.9	0.38	0.47	14.2	0.55	0.61	12.5	0.37	0.41	14.3
ITEM 28	0.43	0.46	13.7	0.45	0.52	13.5	0.35	0.33	14.5	0.31	0.32	14.9	0.46	0.48	13.4	0.33	0.33	14.8
ITEM 29	0.35	0.35	14.5	0.40	0.42	14.0	0.31	0.29	15.0	0.32	0.26	14.9	0.36	0.37	14.4	0.32	0.22	14.8
ITEM 30	0.25	0.28	15.8	0.29	0.34	15.2	0.23	0.17	15.9	0.22	0.05	16.1	0.25	0.32	15.7	0.20	0.09	16.3
COLUMN MEAN	0.63	0.58	11.4	0.67	0.62	11.1	0.56	0.51	12.3	0.54	0.48	12.5	0.66	0.59	11.0	0.52	0.49	12.7
COLUMN S.D.	0.18	0.13	2.2	0.16	0.14	1.9	0.17	0.14	1.9	0.19	0.15	2.1	0.19	0.13	2.4	0.18	0.17	1.9

SAMPLE SIZE	23536	1485	2981	2845	15694	308
POPULATION ESTIMATE	2984583	104503	301603	384751	2120516	43293
COEFFICIENT ALPHA	0.83	0.86	0.81	0.76	0.83	0.79
SPLIT HALF RELIABILITY	0.84	0.87	0.82	0.77	0.84	0.76

	TOTAL		ASIAN		HISPANIC		BLACK		WHITE		AMERICAN INDIAN	
	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.	MEAN	S.D.
FORMULA SCORE	15.1	7.64	16.3	8.10	11.9	7.64	11.2	6.90	16.4	7.31	10.5	7.40
NUMBER RIGHT	18.9	5.53	19.9	5.83	16.7	5.46	16.1	4.93	19.8	5.32	15.7	5.20
NUMBER WRONG	10.8	5.41	9.8	5.67	12.8	5.33	13.4	4.86	9.9	5.23	13.9	5.17
NUMBER OMITS	0.2	0.92	0.2	1.07	0.3	1.27	0.3	1.02	0.2	0.78	0.4	1.42
NUMBER NOT REACHED	0.1	0.89	0.1	0.93	0.2	1.40	0.2	1.36	0.1	0.66	0.1	0.82

Source: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988: Base Year Survey .

Appendix A-4--(continued)

Item Analysis Statistics, History/Citizenship/Geography

	HISPANIC MALE			HISPANIC FEMALE			BLACK MALE			BLACK FEMALE			WHITE MALE			WHITE FEMALE		
	P+	RBIS	DELTA	P+	RBIS	DELTA	P+	RBIS	DELTA	P+	RBIS	DELTA	P+	RBIS	DELTA	P+	RBIS	DELTA
ITEM 1	0.73	0.55	10.5	0.74	0.53	10.4	0.67	0.45	11.2	0.66	0.49	11.4	0.83	0.59	9.2	0.84	0.57	9.1
ITEM 2	0.64	0.64	11.6	0.64	0.55	11.6	0.72	0.61	10.7	0.75	0.56	10.3	0.80	0.70	9.6	0.81	0.63	9.5
ITEM 3	0.82	0.70	9.3	0.86	0.62	8.7	0.80	0.70	9.7	0.84	0.63	9.0	0.91	0.82	7.6	0.94	0.76	6.9
ITEM 4	0.52	0.57	12.8	0.47	0.50	13.3	0.55	0.61	12.5	0.53	0.53	12.7	0.76	0.66	10.1	0.73	0.57	10.6
ITEM 5	0.79	0.63	9.7	0.80	0.61	9.6	0.80	0.54	9.6	0.77	0.58	10.0	0.89	0.64	8.0	0.88	0.69	8.3
ITEM 6	0.71	0.55	10.8	0.78	0.55	9.9	0.77	0.53	10.0	0.78	0.54	9.9	0.86	0.51	8.7	0.86	0.50	8.6
ITEM 7	0.82	0.84	9.4	0.83	0.79	9.2	0.84	0.77	9.1	0.82	0.78	9.3	0.93	0.86	7.0	0.94	0.85	6.6
ITEM 8	0.80	0.72	9.6	0.78	0.67	9.9	0.83	0.68	9.1	0.82	0.67	9.3	0.91	0.73	7.7	0.91	0.70	7.7
ITEM 9	0.81	0.83	9.5	0.81	0.79	9.5	0.84	0.75	9.0	0.84	0.78	9.1	0.94	0.88	6.7	0.95	0.87	6.5
ITEM 10	0.70	0.44	10.9	0.64	0.39	11.5	0.61	0.36	11.9	0.63	0.39	11.6	0.72	0.54	10.7	0.72	0.44	10.7
ITEM 11	0.53	0.56	12.7	0.42	0.49	13.8	0.48	0.48	13.2	0.43	0.40	13.7	0.67	0.69	11.2	0.59	0.62	12.1
ITEM 12	0.44	0.48	13.6	0.51	0.48	12.9	0.42	0.43	13.8	0.49	0.48	13.1	0.54	0.55	12.6	0.60	0.51	12.0
ITEM 13	0.55	0.58	12.5	0.49	0.43	13.1	0.51	0.49	12.9	0.49	0.46	13.1	0.64	0.65	11.6	0.57	0.55	12.3
ITEM 14	0.49	0.42	13.1	0.49	0.44	13.1	0.36	0.36	14.4	0.33	0.33	14.7	0.44	0.44	13.6	0.39	0.43	14.1
ITEM 15	0.40	0.55	14.0	0.39	0.45	14.1	0.34	0.49	14.7	0.33	0.48	14.8	0.51	0.63	12.9	0.50	0.56	13.0
ITEM 16	0.38	0.48	14.2	0.37	0.35	14.3	0.38	0.35	14.2	0.33	0.27	14.7	0.48	0.51	13.2	0.46	0.41	13.4
ITEM 17	0.75	0.69	10.3	0.71	0.61	10.7	0.82	0.66	9.4	0.84	0.54	9.0	0.86	0.68	8.6	0.85	0.60	8.9
ITEM 18	0.71	0.57	10.8	0.69	0.48	11.0	0.68	0.54	11.1	0.69	0.45	11.1	0.81	0.61	9.5	0.81	0.56	9.4
ITEM 19	0.70	0.70	10.9	0.69	0.56	11.0	0.58	0.67	12.2	0.69	0.63	11.1	0.77	0.80	10.0	0.82	0.71	9.3
ITEM 20	0.56	0.52	12.4	0.52	0.53	12.8	0.50	0.50	13.0	0.53	0.46	12.7	0.71	0.62	10.7	0.69	0.58	11.0
ITEM 21	0.67	0.59	11.2	0.48	0.44	13.2	0.55	0.50	12.5	0.42	0.46	13.8	0.78	0.67	10.0	0.63	0.54	11.7
ITEM 22	0.44	0.47	13.6	0.43	0.41	13.7	0.34	0.40	14.7	0.34	0.44	14.7	0.51	0.61	12.9	0.51	0.55	12.9
ITEM 23	0.44	0.51	13.6	0.44	0.38	13.6	0.38	0.39	14.3	0.43	0.39	13.7	0.50	0.53	13.0	0.49	0.46	13.1
ITEM 24	0.48	0.48	13.2	0.46	0.46	13.4	0.42	0.50	13.8	0.48	0.48	13.2	0.57	0.59	12.3	0.56	0.49	12.4
ITEM 25	0.41	0.36	13.9	0.40	0.42	14.0	0.38	0.36	14.2	0.41	0.44	13.9	0.48	0.47	13.2	0.51	0.48	12.9
ITEM 26	0.40	0.44	14.0	0.34	0.37	14.6	0.33	0.34	14.7	0.32	0.29	14.9	0.57	0.55	12.3	0.51	0.51	12.9
ITEM 27	0.42	0.52	13.8	0.40	0.53	14.1	0.36	0.51	14.4	0.41	0.45	13.9	0.57	0.64	12.3	0.54	0.59	12.6
ITEM 28	0.39	0.37	14.2	0.32	0.27	14.9	0.34	0.33	14.6	0.29	0.32	15.2	0.51	0.50	12.9	0.42	0.45	13.8
ITEM 29	0.33	0.24	14.8	0.29	0.33	15.2	0.33	0.19	14.8	0.31	0.33	15.0	0.36	0.36	14.4	0.36	0.39	14.4
ITEM 30	0.24	0.16	15.8	0.22	0.17	16.0	0.23	0.01	15.9	0.21	0.10	16.3	0.27	0.30	15.4	0.23	0.35	16.0
COLUMN MEAN	0.57	0.54	12.2	0.55	0.49	12.4	0.54	0.48	12.5	0.54	0.47	12.5	0.67	0.61	10.9	0.65	0.56	11.1
COLUMN S.D.	0.17	0.15	1.8	0.18	0.13	2.0	0.19	0.16	2.1	0.20	0.14	2.2	0.18	0.13	2.3	0.20	0.12	2.5
SAMPLE SIZE	1428			1532			1372			1454			7785			7797		
POPULATION ESTIMATE	150023			149579			187845			194371			1056913			1051078		
COEFFICIENT ALPHA	0.83			0.78			0.77			0.76			0.84			0.81		
SPLIT HALF RELIABILITY	0.83			0.80			0.77			0.77			0.86			0.82		
	MEAN	S.D.		MEAN	S.D.		MEAN	S.D.		MEAN	S.D.		MEAN	S.D.		MEAN	S.D.	
FORMULA SCORE	12.3	8.03		11.5	7.21		11.2	6.94		11.2	6.86		16.7	7.59		16.0	6.99	
NUMBER RIGHT	17.0	5.72		16.3	5.16		16.1	5.01		16.1	4.85		20.1	5.54		19.6	5.07	
NUMBER WRONG	12.5	5.56		13.1	5.07		13.4	4.91		13.4	4.80		9.7	5.43		10.2	5.02	
NUMBER OMITS	0.3	1.18		0.3	1.34		0.3	1.09		0.3	0.96		0.2	0.72		0.2	0.82	
NUMBER NOT REACHED	0.1	1.18		0.2	1.59		0.2	1.49		0.2	1.23		0.1	0.67		0.1	0.66	

Source: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988: Base Year Survey .

APPENDIX B

DIFFERENTIAL ITEM FUNCTIONING (DIF)

Appendix B-1

Differential Item Functioning (DIF), Reading

MANTEL-HAENSZEL ODDS-RATIO AND OTHER STATISTICS, NUMBER OF TABLES = 21

		NO. LEVELS	LEVEL 1	LEVEL 2
GROUP VARIABLE:	RACE	2	WHITE (REFERENCE)	ASIAN (FOCAL)
RESPONSE VARIABLE:	ITEMSCOR	2	RIGHT	WRONG
STRATIFYING VARIABLE:	# RIGHT	22		

	MH ODDS RATIO	MH CHI-SQUARE	PROB > CHI-SQ	MH D-DIF	STD ERR MH D-DIF	STDZD D-DIF	STD ERR STD D-DIF	REFERENCE N	REFERENCE P+	REFERENCE N0*	FOCAL N	FOCAL P+	FOCAL N0*	IMPACT
ITEM 1	0.82	1.53	0.22	0.47 A	0.36	0.39	0.32	15730	0.96	639	1495	0.96	66	0.00
ITEM 2	1.24	5.82	0.02	-0.51 A	0.21	-0.41	0.18	15724	0.89	639	1494	0.86	66	0.03
ITEM 3	1.28	8.51	0.00	-0.57 A	0.19	-0.44	0.17	15722	0.86	639	1494	0.82	66	0.04
ITEM 4	1.34	20.50	0.00	-0.69 A	0.15	-0.50	0.13	15696	0.65	647	1494	0.58	69	0.07
ITEM 5	1.33	17.29	0.00	-0.66 A	0.16	-0.45	0.13	15657	0.61	647	1485	0.55	69	0.06
ITEM 6	1.02	0.06	0.80	-0.04 A	0.16	-0.03	0.13	15730	0.67	647	1493	0.65	69	0.02
ITEM 7	1.06	0.83	0.36	-0.15 A	0.15	-0.10	0.13	15714	0.47	647	1493	0.45	68	0.02
ITEM 8	0.86	5.29	0.02	0.36 A	0.15	0.25	0.13	15701	0.55	694	1494	0.57	70	-0.01
ITEM 9	0.82	9.20	0.00	0.47 A	0.16	0.39	0.14	15140	0.68	645	1442	0.70	68	-0.03
ITEM 10	0.86	6.32	0.01	0.36 A	0.14	0.31	0.13	15073	0.44	686	1429	0.47	69	-0.03
ITEM 11	0.75	18.60	0.00	0.67 A	0.16	0.50	0.13	15670	0.64	646	1487	0.67	68	-0.03
ITEM 12	1.25	8.62	0.00	-0.52 A	0.18	-0.35	0.14	15675	0.78	646	1488	0.73	68	0.04
ITEM 13	0.90	2.88	0.09	0.25 A	0.15	0.19	0.13	15628	0.56	646	1484	0.57	68	-0.01
ITEM 14	0.85	5.56	0.02	0.38 A	0.16	0.25	0.13	15605	0.54	639	1470	0.55	66	-0.02
ITEM 15	0.90	2.38	0.12	0.25 A	0.16	0.16	0.13	15616	0.52	645	1479	0.53	68	-0.01
ITEM 16	1.01	0.00	0.95	-0.02 A	0.19	-0.01	0.16	15564	0.82	645	1470	0.80	68	0.01
ITEM 17	0.96	0.33	0.57	0.09 A	0.16	0.07	0.13	15521	0.60	645	1469	0.60	68	0.01
ITEM 18	0.93	1.44	0.23	0.18 A	0.15	0.15	0.13	15480	0.58	639	1463	0.59	66	-0.01
ITEM 19	1.06	0.70	0.40	-0.14 A	0.16	-0.10	0.14	15416	0.69	645	1446	0.67	68	0.02
ITEM 20	1.02	0.04	0.84	-0.04 A	0.17	-0.03	0.15	15380	0.76	645	1446	0.74	68	0.01
ITEM 21	1.17	5.19	0.02	-0.37 A	0.16	-0.27	0.13	15348	0.68	639	1444	0.65	66	0.04

Source: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988: Base Year Survey .

Appendix B-1--(continued)

Differential Item Functioning (DIF), Reading

MANTEL-HAENSZEL ODDS-RATIO AND OTHER STATISTICS, NUMBER OF TABLES = 21

		NO. LEVELS	LEVEL 1	LEVEL 2
GROUP VARIABLE:	RACE	2	WHITE (REFERENCE)	HISPANIC (FOCAL)
RESPONSE VARIABLE:	ITEMSCOR	2	RIGHT	WRONG
STRATIFYING VARIABLE:	# RIGHT	22		

	MH ODDS RATIO	MH CHI-SQUARE	PROB > CHI-SQ	MH D-DIF	STD ERR MH D-DIF	STDZD D-DIF	STD ERR STD D-DIF	REFERENCE N	REFERENCE P+	REFERENCE NO*	FOCAL N	FOCAL P+	FOCAL NO*	IMPACT
ITEM 1	0.75	9.73	0.00	0.69 A	0.22	0.57	0.20	15730	0.96	639	2994	0.94	33	0.02
ITEM 2	1.06	0.81	0.37	-0.13 A	0.14	-0.11	0.12	15724	0.89	639	2986	0.80	33	0.08
ITEM 3	1.04	0.49	0.48	-0.10 A	0.13	-0.09	0.11	15722	0.86	639	2988	0.76	33	0.10
ITEM 4	1.12	5.23	0.02	-0.26 A	0.11	-0.21	0.09	15696	0.65	647	2979	0.47	43	0.18
ITEM 5	1.16	9.32	0.00	-0.35 A	0.11	-0.28	0.10	15657	0.61	647	2965	0.43	43	0.19
ITEM 6	1.08	2.43	0.12	-0.17 A	0.11	-0.14	0.09	15730	0.67	639	2993	0.50	33	0.16
ITEM 7	1.14	7.35	0.01	-0.32 A	0.12	-0.25	0.10	15714	0.47	639	2985	0.30	33	0.17
ITEM 8	1.06	1.52	0.22	-0.14 A	0.11	-0.12	0.10	15701	0.55	647	2990	0.38	45	0.18
ITEM 9	0.85	12.42	0.00	0.39 A	0.11	0.32	0.10	15140	0.68	645	2829	0.59	40	0.09
ITEM 10	0.92	3.56	0.06	0.21 A	0.11	0.18	0.10	15073	0.44	644	2817	0.36	40	0.08
ITEM 11	0.75	38.58	0.00	0.68 A	0.11	0.56	0.10	15670	0.64	639	2952	0.54	33	0.09
ITEM 12	1.09	2.94	0.09	-0.21 A	0.12	-0.14	0.10	15675	0.78	646	2952	0.62	43	0.16
ITEM 13	0.93	2.32	0.13	0.16 A	0.11	0.15	0.10	15628	0.56	646	2931	0.44	44	0.12
ITEM 14	0.99	0.03	0.86	0.02 A	0.11	0.03	0.10	15605	0.54	639	2928	0.38	33	0.16
ITEM 15	0.86	9.56	0.00	0.37 A	0.12	0.26	0.10	15616	0.52	645	2915	0.38	43	0.15
ITEM 16	1.02	0.09	0.76	-0.04 A	0.13	-0.02	0.10	15564	0.82	639	2899	0.68	33	0.13
ITEM 17	1.14	7.42	0.01	-0.30 A	0.11	-0.21	0.10	15521	0.60	645	2884	0.42	42	0.19
ITEM 18	0.84	15.37	0.00	0.42 A	0.11	0.35	0.10	15480	0.58	639	2874	0.49	33	0.09
ITEM 19	1.11	4.47	0.03	-0.24 A	0.11	-0.18	0.10	15416	0.69	639	2831	0.53	33	0.17
ITEM 20	0.95	1.16	0.28	0.13 A	0.12	0.10	0.10	15380	0.76	645	2822	0.64	43	0.12
ITEM 21	1.09	3.54	0.06	-0.21 A	0.11	-0.16	0.10	15348	0.68	639	2808	0.53	33	0.16

Source: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988: Base Year Survey .

Appendix B-1--(continued)

Differential Item Functioning (DIF), Reading

MANTEL-HAENSZEL ODDS-RATIO AND OTHER STATISTICS, NUMBER OF TABLES = 21

		NO. LEVELS	LEVEL 1		LEVEL 2	
GROUP VARIABLE:	RACE	2	WHITE	(REFERENCE)	BLACK	(FOCAL)
RESPONSE VARIABLE:	ITEMSCOR	2	RIGHT		WRONG	
STRATIFYING VARIABLE:	# RIGHT	22				

	MH ODDS RATIO	MH CHI-SQUARE	PROB > CHI-SQ	MH D-DIF	STD ERR MH D-DIF	STDZD D-DIF	STD ERR STD D-DIF	REFERENCE N	REFERENCE P+	REFERENCE N0*	FOCAL N	FOCAL P+	FOCAL N0*	IMPACT
ITEM 1	0.70	15.38	0.00	0.85 A	0.22	0.75	0.20	15730	0.96	639	2854	0.93	21	0.03
ITEM 2	1.23	13.06	0.00	-0.49 A	0.14	-0.39	0.12	15724	0.89	639	2842	0.76	21	0.13
ITEM 3	0.96	0.58	0.45	0.10 A	0.13	0.09	0.11	15722	0.86	639	2843	0.75	21	0.12
ITEM 4	1.39	44.67	0.00	-0.78 A	0.12	-0.60	0.10	15696	0.65	647	2837	0.40	30	0.25
ITEM 5	0.77	26.86	0.00	0.60 A	0.12	0.44	0.10	15657	0.61	647	2817	0.47	30	0.15
ITEM 6	1.15	8.88	0.00	-0.34 A	0.11	-0.26	0.10	15730	0.67	639	2845	0.46	21	0.21
ITEM 7	1.09	2.97	0.09	-0.21 A	0.12	-0.17	0.11	15714	0.47	647	2832	0.28	30	0.19
ITEM 8	0.92	2.90	0.09	0.20 A	0.12	0.14	0.10	15701	0.55	647	2832	0.37	29	0.18
ITEM 9	0.78	25.05	0.00	0.58 A	0.12	0.46	0.10	15140	0.68	645	2630	0.57	26	0.10
ITEM 10	0.85	11.61	0.00	0.39 A	0.11	0.36	0.11	15073	0.44	644	2614	0.36	26	0.09
ITEM 11	0.84	12.30	0.00	0.40 A	0.11	0.32	0.10	15670	0.64	639	2805	0.48	21	0.15
ITEM 12	1.29	25.15	0.00	-0.61 A	0.12	-0.40	0.10	15675	0.78	646	2805	0.55	29	0.23
ITEM 13	1.02	0.20	0.65	-0.05 A	0.11	-0.01	0.10	15628	0.56	646	2807	0.40	29	0.16
ITEM 14	0.78	25.94	0.00	0.59 A	0.12	0.47	0.10	15605	0.54	639	2771	0.39	21	0.14
ITEM 15	0.69	48.85	0.00	0.87 A	0.12	0.59	0.10	15616	0.52	645	2730	0.38	27	0.14
ITEM 16	0.86	7.52	0.01	0.36 A	0.13	0.26	0.11	15564	0.82	645	2701	0.68	25	0.14
ITEM 17	0.97	0.44	0.51	0.08 A	0.12	0.09	0.10	15521	0.60	639	2669	0.42	21	0.18
ITEM 18	0.82	17.27	0.00	0.47 A	0.11	0.37	0.10	15480	0.58	639	2642	0.47	21	0.11
ITEM 19	1.26	20.53	0.00	-0.54 A	0.12	-0.41	0.10	15416	0.69	645	2574	0.47	25	0.22
ITEM 20	1.04	0.52	0.47	-0.09 A	0.12	-0.06	0.10	15380	0.76	645	2567	0.59	25	0.16
ITEM 21	1.10	4.06	0.04	-0.23 A	0.12	-0.17	0.10	15348	0.68	639	2544	0.50	21	0.18

Source: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988: Base Year Survey .

Appendix B-1--(continued)

Differential Item Functioning (DIF), Reading

MANTEL-HAENSZEL ODDS-RATIO AND OTHER STATISTICS, NUMBER OF TABLES = 21

		NO. LEVELS	LEVEL 1		LEVEL 2	
GROUP VARIABLE:	RACE	2	WHITE	(REFERENCE)	AM IND	(FOCAL)
RESPONSE VARIABLE:	ITEMSCOR	2	RIGHT		WRONG	
STRATIFYING VARIABLE:	# RIGHT	22				

	MH ODDS RATIO	MH CHI-SQUARE	PROB > CHI-SQ	MH D-DIF		STD ERR MH D-DIF	STDZD D-DIF	STD ERR STD D-DIF	REFERENCE N	REFERENCE P+	REFERENCE NO#	FOCAL N	FOCAL P+	FOCAL NO#	IMPACT
ITEM 1	0.38	11.82	0.00	2.29	C	0.68	2.05	0.65	15730	0.96	639	307	0.95	2	0.00
ITEM 2	1.38	4.86	0.03	-0.77	A	0.34	-0.62	0.31	15724	0.89	647	306	0.73	4	0.16
ITEM 3	0.98	0.00	0.97	0.04	A	0.36	0.02	0.31	15722	0.86	639	306	0.73	2	0.13
ITEM 4	0.88	0.76	0.38	0.30	A	0.32	0.23	0.27	15696	0.65	647	306	0.47	4	0.18
ITEM 5	1.14	0.76	0.38	-0.31	A	0.32	-0.24	0.28	15657	0.61	647	304	0.39	4	0.23
ITEM 6	1.01	0.00	0.98	-0.03	A	0.31	-0.03	0.27	15730	0.67	647	307	0.47	4	0.19
ITEM 7	1.05	0.09	0.77	-0.12	A	0.34	-0.10	0.30	15714	0.47	647	305	0.28	4	0.20
ITEM 8	1.09	0.28	0.60	-0.21	A	0.34	-0.16	0.29	15701	0.55	694	305	0.33	7	0.22
ITEM 9	0.91	0.42	0.52	0.22	A	0.31	0.19	0.29	15140	0.68	645	281	0.54	4	0.14
ITEM 10	0.85	1.20	0.27	0.37	A	0.32	0.33	0.30	15073	0.44	686	279	0.35	7	0.10
ITEM 11	0.74	5.20	0.02	0.72	A	0.31	0.57	0.27	15670	0.64	646	301	0.50	4	0.14
ITEM 12	1.10	0.34	0.56	-0.22	A	0.34	-0.15	0.28	15675	0.78	646	303	0.56	4	0.21
ITEM 13	1.08	0.30	0.59	-0.18	A	0.30	-0.16	0.28	15628	0.56	646	302	0.37	4	0.19
ITEM 14	0.97	0.04	0.85	0.08	A	0.32	0.06	0.29	15605	0.54	639	303	0.34	2	0.19
ITEM 15	0.79	2.35	0.13	0.54	A	0.34	0.40	0.29	15616	0.52	645	298	0.35	4	0.17
ITEM 16	1.06	0.09	0.76	-0.13	A	0.35	-0.09	0.29	15564	0.82	645	297	0.63	4	0.19
ITEM 17	0.90	0.65	0.42	0.26	A	0.31	0.23	0.28	15521	0.60	645	295	0.42	3	0.18
ITEM 18	1.15	0.99	0.32	-0.34	A	0.31	-0.29	0.29	15480	0.58	639	295	0.39	2	0.19
ITEM 19	1.20	1.59	0.21	-0.42	A	0.32	-0.32	0.28	15416	0.69	645	297	0.46	4	0.23
ITEM 20	0.99	0.00	0.97	0.03	A	0.32	0.03	0.28	15380	0.76	645	295	0.58	4	0.17
ITEM 21	1.23	2.34	0.13	-0.49	A	0.31	-0.40	0.28	15348	0.68	639	295	0.46	2	0.22

Source: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988: Base Year Survey .

Appendix B-1--(continued)

Differential Item Functioning (DIF), Reading

MANTEL-HAENSZEL ODDS-RATIO AND OTHER STATISTICS, NUMBER OF TABLES = 21

		NO. LEVELS	LEVEL 1	LEVEL 2
GROUP VARIABLE:	SEX	2	MALE (REFERENCE)	FEMALE (FOCAL)
RESPONSE VARIABLE:	ITEMSCOR	2	RIGHT	WRONG
STRATIFYING VARIABLE:	# RIGHT	22		

	MH ODDS RATIO	MH CHI-SQUARE	PROB > CHI-SQ	MH D-DIF		STD ERR MH D-DIF	STDZD D-DIF	STD ERR STD D-DIF	REFERENCE N	REFERENCE P+	REFERENCE N0*	FOCAL N	FOCAL P+	FOCAL N0*	IMPACT
ITEM 1	0.75	17.86	0.00	0.68	A	0.16	0.57	0.15	11639	0.94	320	11791	0.96	436	-0.02
ITEM 2	1.21	20.08	0.00	-0.45	A	0.10	-0.36	0.09	11628	0.85	320	11776	0.86	436	-0.01
ITEM 3	0.93	2.98	0.08	0.16	A	0.09	0.15	0.08	11629	0.81	320	11774	0.85	436	-0.04
ITEM 4	0.84	32.27	0.00	0.42	A	0.07	0.31	0.06	11609	0.55	339	11752	0.63	451	-0.08
ITEM 5	1.05	2.71	0.10	-0.12	A	0.07	-0.07	0.06	11566	0.55	337	11710	0.58	451	-0.04
ITEM 6	1.53	178.73	0.00	-1.00	B	0.08	-0.76	0.06	11640	0.63	320	11776	0.60	436	0.03
ITEM 7	1.14	18.06	0.00	-0.31	A	0.07	-0.23	0.06	11632	0.41	320	11747	0.43	436	-0.02
ITEM 8	1.21	37.19	0.00	-0.45	A	0.07	-0.31	0.06	11614	0.50	338	11756	0.51	450	-0.02
ITEM 9	0.77	70.71	0.00	0.62	A	0.07	0.52	0.06	11005	0.61	329	11363	0.69	449	-0.08
ITEM 10	1.11	11.86	0.00	-0.24	A	0.07	-0.23	0.06	10959	0.42	329	11297	0.43	448	-0.01
ITEM 11	0.82	38.46	0.00	0.46	A	0.07	0.34	0.06	11547	0.56	320	11717	0.65	436	-0.08
ITEM 12	0.70	98.83	0.00	0.85	A	0.09	0.57	0.06	11544	0.68	336	11727	0.77	449	-0.10
ITEM 13	1.42	134.74	0.00	-0.82	A	0.07	-0.66	0.06	11508	0.54	337	11691	0.51	449	0.03
ITEM 14	1.00	0.00	0.99	0.00	A	0.07	0.00	0.06	11482	0.47	320	11638	0.52	436	-0.05
ITEM 15	0.96	1.39	0.24	0.09	A	0.08	0.06	0.06	11436	0.46	334	11640	0.51	448	-0.06
ITEM 16	0.99	0.02	0.89	0.01	A	0.09	0.00	0.07	11371	0.76	320	11598	0.80	436	-0.05
ITEM 17	0.88	17.18	0.00	0.31	A	0.07	0.23	0.06	11322	0.52	320	11553	0.59	436	-0.07
ITEM 18	1.00	0.01	0.91	-0.01	A	0.07	0.01	0.06	11252	0.53	320	11539	0.57	436	-0.04
ITEM 19	0.97	0.88	0.35	0.07	A	0.08	0.06	0.06	11157	0.61	320	11438	0.67	436	-0.05
ITEM 20	0.83	29.49	0.00	0.44	A	0.08	0.34	0.07	11143	0.68	333	11402	0.75	447	-0.07
ITEM 21	1.06	3.78	0.05	-0.15	A	0.08	-0.10	0.06	11105	0.62	320	11365	0.65	436	-0.03

Source: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988: Base Year Survey .

Appendix B-2

Differential Item Functioning (DIF), Mathematics

MANTEL-HAENSZEL ODDS-RATIO AND OTHER STATISTICS, NUMBER OF TABLES = 40

		NO. LEVELS	LEVEL 1	LEVEL 2
GROUP VARIABLE:	RACE	2	WHITE (REFERENCE)	ASIAN (FOCAL)
RESPONSE VARIABLE:	ITEMSCOR	2	RIGHT	WRONG
STRATIFYING VARIABLE:	# RIGHT	41		

	MH ODDS RATIO	MH CHI-SQUARE	PROB > CHI-SQ	MH D-DIF		STD ERR MH D-DIF	STDZD D-DIF	STD ERR STD D-DIF	REFERENCE N	REFERENCE P+	REFERENCE NO*	FOCAL N	FOCAL P+	FOCAL NO*	IMPACT
ITEM 1	0.94	0.36	0.55	0.13	A	0.21	0.10	0.17	15145	0.80	110	1451	0.83	26	-0.04
ITEM 2	1.13	3.72	0.05	-0.29	A	0.15	-0.21	0.13	15656	0.57	98	1483	0.60	26	-0.03
ITEM 3	1.04	0.45	0.50	-0.10	A	0.14	-0.08	0.13	15423	0.51	99	1456	0.54	26	-0.03
ITEM 4	0.90	2.26	0.13	0.24	A	0.16	0.17	0.13	15614	0.56	113	1476	0.64	29	-0.08
ITEM 5	0.84	5.67	0.02	0.41	A	0.17	0.28	0.14	15338	0.59	100	1454	0.67	28	-0.08
ITEM 6	1.13	3.52	0.06	-0.28	A	0.15	-0.22	0.13	15467	0.50	129	1457	0.53	33	-0.03
ITEM 7	1.03	0.17	0.68	-0.07	A	0.16	-0.04	0.13	15572	0.49	108	1474	0.55	27	-0.06
ITEM 8	0.64	49.18	0.00	1.06	B	0.15	0.77	0.13	15692	0.42	98	1487	0.56	26	-0.14
ITEM 9	0.81	10.14	0.00	0.49	A	0.15	0.35	0.13	15617	0.49	101	1485	0.59	28	-0.10
ITEM 10	1.70	65.73	0.00	-1.24	B	0.16	-0.88	0.13	15639	0.48	102	1477	0.45	28	0.03
ITEM 11	0.94	0.83	0.36	0.14	A	0.15	0.10	0.13	15573	0.41	107	1473	0.48	29	-0.07
ITEM 12	0.74	19.85	0.00	0.70	A	0.16	0.48	0.13	15632	0.50	99	1486	0.61	28	-0.11
ITEM 13	0.84	5.49	0.02	0.41	A	0.17	0.26	0.14	15483	0.60	107	1467	0.68	29	-0.08
ITEM 14	0.83	7.59	0.01	0.45	A	0.16	0.30	0.13	15544	0.57	106	1473	0.66	29	-0.09
ITEM 15	0.83	5.48	0.02	0.43	A	0.18	0.35	0.16	15426	0.76	94	1469	0.81	26	-0.06
ITEM 16	0.95	0.39	0.53	0.13	A	0.19	0.11	0.18	15655	0.82	95	1486	0.85	26	-0.03
ITEM 17	0.97	0.11	0.75	0.06	A	0.17	0.05	0.15	15639	0.75	95	1484	0.78	26	-0.04
ITEM 18	0.84	6.31	0.01	0.42	A	0.16	0.28	0.13	15571	0.58	105	1481	0.67	29	-0.09
ITEM 19	0.66	16.88	0.00	0.96	A	0.23	0.77	0.21	15463	0.84	96	1477	0.90	27	-0.06
ITEM 20	0.84	1.74	0.19	0.40	A	0.29	0.31	0.25	14215	0.91	623	1399	0.93	133	-0.02
ITEM 21	1.47	30.10	0.00	-0.90	A	0.16	-0.72	0.15	15559	0.76	94	1453	0.73	27	0.02
ITEM 22	0.92	0.91	0.34	0.19	A	0.19	0.13	0.16	15551	0.77	103	1473	0.81	29	-0.04
ITEM 23	1.16	5.29	0.02	-0.35	A	0.15	-0.31	0.14	15512	0.70	96	1467	0.70	27	0.00
ITEM 24	1.12	2.85	0.09	-0.26	A	0.15	-0.21	0.14	15603	0.65	102	1475	0.67	28	-0.02
ITEM 25	1.39	20.88	0.00	-0.77	A	0.17	-0.58	0.14	15656	0.74	104	1482	0.73	29	0.01
ITEM 26	1.24	9.22	0.00	-0.51	A	0.17	-0.37	0.14	15533	0.69	100	1467	0.70	28	-0.01
ITEM 27	1.20	5.24	0.02	-0.42	A	0.18	-0.25	0.14	15548	0.69	100	1474	0.72	28	-0.03
ITEM 28	1.36	23.58	0.00	-0.73	A	0.15	-0.57	0.13	15643	0.64	96	1483	0.63	27	0.01
ITEM 29	0.85	5.19	0.02	0.38	A	0.16	0.26	0.13	15537	0.58	106	1471	0.66	27	-0.08
ITEM 30	1.03	0.15	0.69	-0.06	A	0.15	-0.05	0.13	15343	0.57	114	1446	0.62	29	-0.05
ITEM 31	0.93	0.90	0.34	0.17	A	0.17	0.11	0.14	15429	0.68	102	1465	0.73	27	-0.06
ITEM 32	1.08	1.17	0.28	-0.19	A	0.17	-0.15	0.15	15428	0.72	106	1452	0.75	27	-0.03
ITEM 33	1.09	2.28	0.13	-0.21	A	0.14	-0.18	0.13	15591	0.50	102	1475	0.52	28	-0.02
ITEM 34	0.63	39.72	0.00	1.07	B	0.17	0.75	0.14	15250	0.58	94	1443	0.71	26	-0.13
ITEM 35	1.24	12.00	0.00	-0.50	A	0.15	-0.42	0.13	15425	0.63	95	1462	0.63	26	0.00
ITEM 36	0.95	0.55	0.46	0.12	A	0.16	0.08	0.13	15493	0.47	102	1463	0.55	28	-0.08
ITEM 37	1.89	94.36	0.00	-1.50	C	0.16	-1.03	0.13	15564	0.54	133	1471	0.49	31	0.05
ITEM 38	1.01	0.01	0.92	-0.02	A	0.14	-0.02	0.13	15411	0.45	95	1452	0.49	26	-0.04
ITEM 39	0.86	4.12	0.04	0.35	A	0.17	0.20	0.13	15444	0.46	123	1466	0.56	29	-0.10
ITEM 40	0.61	52.30	0.00	1.16	B	0.16	0.75	0.13	15190	0.34	111	1448	0.49	29	-0.16

Source: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988: Base Year Survey .

Appendix B-2--(continued)

Differential Item Functioning (DIF), Mathematics

MANTEL-HAENSZEL ODDS-RATIO AND OTHER STATISTICS, NUMBER OF TABLES = 40

		NO. LEVELS	LEVEL 1	LEVEL 2
GROUP VARIABLE:	RACE	2	WHITE (REFERENCE)	HISPANIC (FOCAL)
RESPONSE VARIABLE:	ITEMSCOR	2	RIGHT	WRONG
STRATIFYING VARIABLE:	B RIGHT	41		

	MH ODDS RATIO	MH CHI-SQUARE	PROB > CHI-SQ	MH D-DIF	STD ERR MH D-DIF	STDZD D-DIF	STD ERR STD D-DIF	REFERENCE N	REFERENCE P+	REFERENCE NO*	FOCAL N	FOCAL P+	FOCAL NO*	IMPACT
ITEM 1	1.14	5.98	0.01	-0.30 A	0.12	-0.22	0.10	15145	0.80	102	2849	0.63	11	0.17
ITEM 2	1.09	3.23	0.07	-0.20 A	0.11	-0.17	0.10	15656	0.57	94	2940	0.40	8	0.17
ITEM 3	0.99	0.08	0.77	0.03 A	0.10	0.03	0.10	15423	0.51	94	2854	0.44	7	0.06
ITEM 4	1.11	4.84	0.03	-0.25 A	0.11	-0.20	0.10	15614	0.56	113	2927	0.37	23	0.19
ITEM 5	0.93	2.28	0.13	0.17 A	0.11	0.13	0.10	15338	0.59	100	2869	0.43	13	0.15
ITEM 6	0.99	0.04	0.84	0.02 A	0.10	0.03	0.10	15467	0.50	97	2889	0.39	8	0.11
ITEM 7	1.13	5.23	0.02	-0.28 A	0.12	-0.22	0.10	15572	0.49	113	2901	0.29	20	0.19
ITEM 8	0.93	2.02	0.16	0.17 A	0.12	0.17	0.11	15692	0.42	98	2955	0.29	8	0.13
ITEM 9	0.92	3.08	0.08	0.19 A	0.11	0.19	0.10	15617	0.49	97	2941	0.37	8	0.11
ITEM 10	1.24	19.64	0.00	-0.51 A	0.11	-0.41	0.10	15639	0.48	102	2937	0.30	10	0.19
ITEM 11	0.92	2.83	0.09	0.20 A	0.11	0.20	0.11	15573	0.41	97	2911	0.29	10	0.12
ITEM 12	0.76	32.69	0.00	0.66 A	0.11	0.50	0.10	15632	0.50	99	2935	0.38	11	0.12
ITEM 13	0.94	1.32	0.25	0.14 A	0.12	0.10	0.10	15483	0.60	103	2886	0.42	17	0.17
ITEM 14	0.94	1.47	0.22	0.14 A	0.12	0.11	0.10	15544	0.57	94	2899	0.40	7	0.17
ITEM 15	0.80	20.18	0.00	0.52 A	0.11	0.45	0.11	15426	0.76	94	2880	0.68	7	0.07
ITEM 16	0.82	14.40	0.00	0.48 A	0.12	0.40	0.11	15655	0.82	95	2955	0.75	8	0.06
ITEM 17	1.03	0.35	0.55	-0.07 A	0.11	-0.06	0.10	15639	0.75	95	2947	0.64	8	0.11
ITEM 18	0.91	4.40	0.04	0.23 A	0.11	0.21	0.10	15571	0.58	331	2939	0.43	20	0.15
ITEM 19	0.90	3.66	0.06	0.25 A	0.13	0.19	0.11	15463	0.84	96	2911	0.74	8	0.10
ITEM 20	0.93	1.02	0.31	0.16 A	0.16	0.14	0.13	14215	0.91	623	2739	0.82	29	0.09
ITEM 21	1.27	24.77	0.00	-0.55 A	0.11	-0.46	0.10	15559	0.76	320	2930	0.58	17	0.17
ITEM 22	1.23	16.80	0.00	-0.49 A	0.12	-0.33	0.10	15551	0.77	103	2937	0.57	16	0.20
ITEM 23	1.12	5.54	0.02	-0.26 A	0.11	-0.22	0.10	15512	0.70	94	2940	0.58	7	0.13
ITEM 24	1.00	0.00	1.00	0.00 A	0.11	-0.03	0.10	15603	0.65	95	2946	0.51	8	0.14
ITEM 25	1.40	50.67	0.00	-0.80 A	0.11	-0.62	0.10	15656	0.74	104	2959	0.53	16	0.21
ITEM 26	0.85	11.99	0.00	0.38 A	0.11	0.33	0.10	15533	0.69	94	2898	0.58	7	0.11
ITEM 27	1.05	1.03	0.31	-0.12 A	0.12	-0.07	0.10	15548	0.69	100	2911	0.49	13	0.20
ITEM 28	1.26	25.45	0.00	-0.54 A	0.11	-0.44	0.10	15643	0.64	96	2942	0.44	8	0.20
ITEM 29	0.76	31.94	0.00	0.63 A	0.11	0.49	0.10	15537	0.58	95	2913	0.46	7	0.12
ITEM 30	0.92	3.60	0.06	0.21 A	0.11	0.17	0.10	15343	0.57	101	2857	0.45	9	0.12
ITEM 31	1.07	1.84	0.18	-0.16 A	0.11	-0.12	0.10	15429	0.68	102	2883	0.49	14	0.18
ITEM 32	0.89	5.98	0.01	0.27 A	0.11	0.25	0.10	15428	0.72	320	2881	0.62	15	0.10
ITEM 33	1.05	1.22	0.27	-0.11 A	0.10	-0.13	0.10	15591	0.50	94	2926	0.41	7	0.08
ITEM 34	0.86	10.13	0.00	0.35 A	0.11	0.29	0.10	15250	0.58	94	2814	0.46	8	0.13
ITEM 35	1.11	5.61	0.02	-0.25 A	0.10	-0.21	0.10	15425	0.63	94	2860	0.48	6	0.15
ITEM 36	1.01	0.02	0.88	-0.02 A	0.12	-0.03	0.10	15493	0.47	102	2886	0.31	13	0.16
ITEM 37	1.41	43.81	0.00	-0.81 A	0.12	-0.60	0.10	15564	0.54	133	2906	0.30	39	0.23
ITEM 38	0.89	6.64	0.01	0.27 A	0.10	0.25	0.10	15411	0.45	95	2842	0.40	9	0.05
ITEM 39	1.10	3.06	0.08	-0.22 A	0.13	-0.18	0.11	15444	0.46	95	2884	0.27	8	0.19
ITEM 40	0.96	0.53	0.47	0.10 A	0.13	0.07	0.12	15190	0.34	100	2804	0.21	11	0.13

Source: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988: Base Year Survey .

Appendix B-2--(continued)

Differential Item Functioning (DIF), Mathematics

MANTEL-HAENSZEL ODDS-RATIO AND OTHER STATISTICS, NUMBER OF TABLES = 40

		NO. LEVELS	LEVEL 1		LEVEL 2	
GROUP VARIABLE:	RACE	2	WHITE	(REFERENCE)	BLACK	(FOCAL)
RESPONSE VARIABLE:	ITEMSCOR	2	RIGHT		WRONG	
STRATIFYING VARIABLE:	# RIGHT	41				

	MH ODDS RATIO	MH CHI-SQUARE	PROB > CHI-SQ	MH D-DIF	STD ERR MH D-DIF	STDZD D-DIF	STD ERR STD D-DIF	REFERENCE N	REFERENCE P+	REFERENCE NO#	FOCAL N	FOCAL P+	FOCAL NO#	IMPACT
ITEM 1	1.06	1.00	0.32	-0.13 A	0.12	-0.10	0.10	15145	0.80	97	2734	0.58	1	0.22
ITEM 2	1.25	20.22	0.00	-0.52 A	0.12	-0.45	0.11	15656	0.57	97	2801	0.33	1	0.23
ITEM 3	0.83	17.14	0.00	0.44 A	0.11	0.40	0.10	15423	0.51	97	2707	0.47	1	0.04
ITEM 4	0.81	18.23	0.00	0.50 A	0.12	0.39	0.10	15614	0.56	97	2794	0.39	3	0.17
ITEM 5	0.88	5.75	0.02	0.29 A	0.12	0.19	0.10	15338	0.59	97	2709	0.40	3	0.19
ITEM 6	0.90	5.46	0.02	0.26 A	0.11	0.22	0.11	15467	0.50	97	2736	0.39	3	0.11
ITEM 7	0.85	8.75	0.00	0.38 A	0.13	0.27	0.11	15572	0.49	102	2777	0.30	6	0.19
ITEM 8	1.02	0.16	0.69	-0.05 A	0.13	-0.04	0.12	15692	0.42	98	2828	0.24	3	0.17
ITEM 9	0.90	4.75	0.03	0.25 A	0.11	0.25	0.11	15617	0.49	96	2787	0.35	1	0.14
ITEM 10	1.15	7.34	0.01	-0.33 A	0.12	-0.29	0.11	15639	0.48	97	2793	0.27	3	0.21
ITEM 11	1.07	1.65	0.20	-0.16 A	0.12	-0.09	0.12	15573	0.41	104	2765	0.24	13	0.17
ITEM 12	0.75	32.67	0.00	0.69 A	0.12	0.53	0.11	15632	0.50	99	2797	0.33	7	0.16
ITEM 13	0.82	16.08	0.00	0.48 A	0.12	0.37	0.10	15483	0.60	107	2758	0.40	14	0.20
ITEM 14	0.76	30.37	0.00	0.66 A	0.12	0.49	0.10	15544	0.57	98	2765	0.39	5	0.18
ITEM 15	0.64	73.47	0.00	1.03 B	0.12	0.88	0.11	15426	0.76	94	2718	0.69	2	0.07
ITEM 16	0.68	49.35	0.00	0.90 A	0.13	0.78	0.12	15655	0.82	95	2807	0.75	3	0.06
ITEM 17	0.92	3.22	0.07	0.20 A	0.11	0.20	0.11	15639	0.75	94	2805	0.63	1	0.11
ITEM 18	0.88	6.59	0.01	0.30 A	0.12	0.24	0.10	15571	0.58	331	2783	0.39	13	0.19
ITEM 19	0.58	89.24	0.00	1.33 B	0.14	1.08	0.12	15463	0.84	94	2747	0.78	2	0.06
ITEM 20	0.74	19.34	0.00	0.70 A	0.16	0.57	0.14	14215	0.91	623	2559	0.81	14	0.09
ITEM 21	1.23	18.76	0.00	-0.49 A	0.11	-0.39	0.10	15559	0.76	320	2778	0.55	8	0.21
ITEM 22	1.44	50.31	0.00	-0.86 A	0.12	-0.58	0.10	15551	0.77	95	2818	0.48	6	0.29
ITEM 23	1.06	1.20	0.27	-0.13 A	0.11	-0.13	0.10	15512	0.70	96	2794	0.54	5	0.16
ITEM 24	1.07	1.77	0.18	-0.15 A	0.11	-0.15	0.10	15603	0.65	95	2797	0.46	2	0.19
ITEM 25	2.74	429.08	0.00	-2.37 C	0.12	-1.87	0.10	15656	0.74	96	2815	0.36	5	0.39
ITEM 26	0.96	0.79	0.38	0.10 A	0.11	0.09	0.10	15533	0.69	100	2755	0.51	5	0.18
ITEM 27	1.17	8.60	0.00	-0.36 A	0.12	-0.25	0.10	15548	0.69	94	2757	0.41	2	0.28
ITEM 28	1.36	41.15	0.00	-0.72 A	0.11	-0.56	0.10	15643	0.64	94	2811	0.38	2	0.26
ITEM 29	0.82	15.11	0.00	0.46 A	0.12	0.35	0.10	15537	0.58	95	2758	0.40	2	0.19
ITEM 30	0.73	42.65	0.00	0.74 A	0.11	0.61	0.10	15343	0.57	103	2666	0.46	9	0.11
ITEM 31	1.39	44.06	0.00	-0.78 A	0.12	-0.59	0.10	15429	0.68	102	2704	0.39	8	0.28
ITEM 32	1.02	0.24	0.63	-0.06 A	0.11	-0.01	0.10	15428	0.72	320	2698	0.56	8	0.17
ITEM 33	1.10	4.47	0.03	-0.23 A	0.11	-0.23	0.11	15591	0.50	95	2751	0.38	5	0.11
ITEM 34	0.81	17.41	0.00	0.49 A	0.12	0.35	0.10	15250	0.58	94	2644	0.42	4	0.16
ITEM 35	1.00	0.00	0.99	0.00 A	0.11	0.02	0.10	15425	0.63	94	2668	0.47	1	0.16
ITEM 36	1.16	7.76	0.01	-0.35 A	0.13	-0.28	0.11	15493	0.47	95	2704	0.26	5	0.21
ITEM 37	1.87	120.40	0.00	-1.47 B	0.14	-1.10	0.12	15564	0.54	97	2702	0.22	4	0.32
ITEM 38	0.83	16.74	0.00	0.45 A	0.11	0.42	0.11	15411	0.45	95	2629	0.40	4	0.05
ITEM 39	1.05	0.65	0.42	-0.11 A	0.13	-0.12	0.12	15444	0.46	95	2665	0.25	4	0.21
ITEM 40	0.96	0.36	0.55	0.09 A	0.14	0.01	0.13	15190	0.34	111	2577	0.19	13	0.15

Source: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988: Base Year Survey .

Appendix B-2--(continued)

Differential Item Functioning (DIF), Mathematics

MANTEL-HAENSZEL ODDS-RATIO AND OTHER STATISTICS, NUMBER OF TABLES = 40

		NO. LEVELS	LEVEL 1		LEVEL 2	
GROUP VARIABLE:	RACE	2	WHITE	(REFERENCE)	AM IND	(FOCAL)
RESPONSE VARIABLE:	ITEMSCOR	2	RIGHT		WRONG	
STRATIFYING VARIABLE:	# RIGHT	41				

	MH ODDS RATIO	MH CHI-SQUARE	PROB > CHI-SQ	MH D-DIF		STD ERR MH D-DIF	STDZD D-DIF	STD ERR STD D-DIF	REFERENCE N	REFERENCE P+	REFERENCE NO*	FOCAL N	FOCAL P+	FOCAL NO*	IMPACT
ITEM 1	0.96	0.06	0.81	0.10	A	0.33	0.08	0.29	15145	0.80	534	295	0.62	0	0.18
ITEM 2	1.44	7.61	0.01	-0.85	A	0.32	-0.72	0.29	15656	0.57	544	304	0.32	0	0.25
ITEM 3	0.91	0.51	0.47	0.22	A	0.28	0.23	0.28	15423	0.51	541	296	0.45	0	0.06
ITEM 4	0.99	0.00	0.99	0.02	A	0.31	0.01	0.29	15614	0.56	538	303	0.36	0	0.20
ITEM 5	0.93	0.24	0.62	0.17	A	0.31	0.14	0.28	15338	0.59	535	299	0.40	0	0.19
ITEM 6	1.02	0.01	0.93	-0.05	A	0.30	-0.04	0.29	15467	0.50	564	297	0.37	2	0.13
ITEM 7	1.04	0.05	0.82	-0.10	A	0.34	-0.07	0.31	15572	0.49	538	299	0.28	0	0.21
ITEM 8	1.02	0.01	0.93	-0.05	A	0.34	-0.05	0.32	15692	0.42	540	303	0.25	0	0.17
ITEM 9	0.86	1.51	0.22	0.37	A	0.29	0.36	0.28	15617	0.49	537	302	0.37	0	0.12
ITEM 10	1.11	0.48	0.49	-0.24	A	0.32	-0.22	0.30	15639	0.48	538	300	0.29	0	0.19
ITEM 11	0.77	3.56	0.06	0.62	A	0.32	0.56	0.30	15573	0.41	532	298	0.30	0	0.11
ITEM 12	0.84	1.33	0.25	0.42	A	0.34	0.31	0.29	15632	0.50	532	302	0.32	0	0.17
ITEM 13	1.01	0.00	0.98	-0.03	A	0.32	-0.02	0.28	15483	0.60	531	298	0.38	0	0.22
ITEM 14	0.97	0.02	0.89	0.07	A	0.33	0.05	0.29	15544	0.57	531	296	0.36	0	0.21
ITEM 15	0.74	4.87	0.03	0.71	A	0.31	0.64	0.30	15426	0.76	530	296	0.68	0	0.08
ITEM 16	0.86	0.99	0.32	0.36	A	0.34	0.31	0.31	15655	0.82	534	300	0.73	0	0.09
ITEM 17	1.02	0.01	0.94	-0.04	A	0.30	-0.03	0.29	15639	0.75	534	300	0.62	0	0.12
ITEM 18	0.87	0.98	0.32	0.32	A	0.31	0.27	0.28	15571	0.58	765	301	0.41	1	0.17
ITEM 19	0.72	4.68	0.03	0.78	A	0.35	0.66	0.32	15463	0.8[illegible]	538	300	0.76	0	0.08
ITEM 20	1.06	0.08	0.78	-0.14	A	0.39	-0.10	0.34	14215	0.91	1515	277	0.77	2	0.14
ITEM 21	1.31	3.84	0.05	-0.63	A	0.31	-0.52	0.28	15559	0.76	768	304	0.55	1	0.21
ITEM 22	1.10	0.35	0.55	-0.22	A	0.33	-0.16	0.27	15551	0.77	540	304	0.55	0	0.22
ITEM 23	1.05	0.11	0.74	-0.12	A	0.30	-0.10	0.28	15512	0.70	549	304	0.56	0	0.14
ITEM 24	1.01	0.00	0.96	-0.03	A	0.30	-0.03	0.27	15603	0.65	542	304	0.48	0	0.17
ITEM 25	1.46	7.94	0.00	-0.90	A	0.31	-0.71	0.27	15656	0.74	544	305	0.49	0	0.25
ITEM 26	0.94	0.16	0.69	0.14	A	0.30	0.12	0.28	15533	0.69	536	294	0.52	0	0.16
ITEM 27	1.33	3.61	0.06	-0.67	A	0.34	-0.46	0.28	15548	0.69	538	298	0.41	0	0.28
ITEM 28	1.07	0.23	0.63	-0.17	A	0.31	-0.14	0.28	15643	0.64	541	304	0.44	0	0.20
ITEM 29	0.86	1.03	0.31	0.35	A	0.32	0.27	0.28	15537	0.58	537	293	0.41	0	0.18
ITEM 30	0.85	1.39	0.24	0.38	A	0.30	0.33	0.28	15343	0.57	536	290	0.45	0	0.13
ITEM 31	1.04	0.04	0.84	-0.09	A	0.32	-0.07	0.28	15429	0.68	536	290	0.47	0	0.21
ITEM 32	1.04	0.08	0.78	-0.10	A	0.30	-0.09	0.28	15428	0.72	762	294	0.56	1	0.16
ITEM 33	0.99	0.00	0.96	0.03	A	0.29	0.03	0.28	15591	0.50	539	299	0.42	0	0.08
ITEM 34	0.79	2.99	0.08	0.56	A	0.31	0.46	0.28	15250	0.58	533	291	0.44	0	0.14
ITEM 35	1.06	0.30	0.58	-0.15	A	0.30	-0.13	0.28	15425	0.63	538	292	0.46	0	0.16
ITEM 36	1.11	0.45	0.50	-0.24	A	0.33	-0.21	0.31	15493	0.47	536	292	0.27	0	0.20
ITEM 37	1.33	3.38	0.07	-0.67	A	0.35	-0.51	0.31	15564	0.54	572	300	0.28	2	0.26
ITEM 38	1.01	0.00	0.97	-0.03	A	0.30	-0.03	0.29	15411	0.45	539	291	0.36	0	0.09
ITEM 39	1.12	0.41	0.52	-0.26	A	0.36	-0.22	0.32	15444	0.46	565	298	0.24	2	0.21
ITEM 40	0.78	2.16	0.14	0.57	A	0.37	0.50	0.34	15190	0.34	533	284	0.22	0	0.12

Source: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988: Base Year Survey .

Appendix B-2--(continued)

Differential Item Functioning (DIF), Mathematics

MANTEL-HAENSZEL ODDS-RATIO AND OTHER STATISTICS, NUMBER OF TABLES = 40

		NO. LEVELS	LEVEL 1		LEVEL 2	
GROUP VARIABLE:	SEX	2	MALE	(REFERENCE)	FEMALE	(FOCAL)
RESPONSE VARIABLE:	ITEMSCOR	2	RIGHT		WRONG	
STRATIFYING VARIABLE:	# RIGHT	41				

	MH ODDS RATIO	MH CHI-SQUARE	PROB > CHI-SQ	MH D-DIF		STD ERR MH D-DIF	STDZD D-DIF	STD ERR STD D-DIF	REFERENCE N	REFERENCE P+	REFERENCE NO*	FOCAL N	FOCAL P+	FOCAL NO*	IMPACT
ITEM 1	0.73	70.80	0.00	0.74	A	0.09	0.51	0.07	11168	0.73	92	11349	0.76	36	-0.03
ITEM 2	1.12	14.37	0.00	-0.27	A	0.07	-0.21	0.06	11546	0.53	95	11685	0.50	38	0.04
ITEM 3	0.87	26.63	0.00	0.33	A	0.06	0.31	0.06	11340	0.49	92	11441	0.51	36	-0.02
ITEM 4	0.87	19.62	0.00	0.32	A	0.07	0.24	0.06	11507	0.51	97	11653	0.52	36	-0.01
ITEM 5	1.05	2.26	0.13	-0.11	A	0.07	-0.08	0.06	11294	0.56	98	11425	0.54	37	0.02
ITEM 6	0.98	0.59	0.44	0.05	A	0.07	0.05	0.06	11390	0.48	96	11499	0.47	37	0.01
ITEM 7	0.86	21.70	0.00	0.35	A	0.08	0.25	0.06	11464	0.44	97	11604	0.44	38	-0.01
ITEM 8	0.90	11.50	0.00	0.25	A	0.07	0.19	0.06	11582	0.39	96	11731	0.39	39	0.00
ITEM 9	1.19	34.39	0.00	-0.41	A	0.07	-0.33	0.06	11526	0.49	92	11647	0.44	37	0.05
ITEM 10	0.90	11.11	0.00	0.24	A	0.07	0.19	0.06	11526	0.43	96	11664	0.43	38	0.00
ITEM 11	1.16	23.92	0.00	-0.35	A	0.07	-0.29	0.06	11457	0.40	94	11604	0.35	36	0.05
ITEM 12	0.71	118.11	0.00	0.80	A	0.07	0.58	0.06	11525	0.44	102	11670	0.49	43	-0.04
ITEM 13	1.13	14.89	0.00	-0.29	A	0.08	-0.21	0.06	11418	0.57	94	11520	0.54	37	0.04
ITEM 14	1.28	58.59	0.00	-0.57	A	0.08	-0.40	0.06	11460	0.56	93	11560	0.50	37	0.06
ITEM 15	0.84	29.29	0.00	0.42	A	0.08	0.36	0.07	11350	0.73	93	11480	0.75	36	-0.02
ITEM 16	1.00	0.00	0.99	0.00	A	0.08	0.01	0.07	11551	0.80	95	11699	0.80	37	0.01
ITEM 17	0.96	1.65	0.20	0.10	A	0.07	0.09	0.07	11534	0.72	92	11689	0.72	36	0.00
ITEM 18	0.93	5.00	0.03	0.16	A	0.07	0.12	0.06	11478	0.54	107	11645	0.54	50	0.00
ITEM 19	0.77	48.41	0.00	0.63	A	0.09	0.51	0.08	11374	0.81	98	11570	0.83	36	-0.03
ITEM 20	0.55	147.52	0.00	1.41	B	0.12	1.08	0.10	10400	0.86	476	10825	0.90	326	-0.04
ITEM 21	1.58	197.61	0.00	-1.08	B	0.08	-0.87	0.06	11467	0.75	93	11607	0.66	37	0.09
ITEM 22	1.24	37.29	0.00	-0.51	A	0.08	-0.34	0.06	11477	0.72	99	11646	0.69	40	0.04
ITEM 23	1.05	2.90	0.09	-0.12	A	0.07	-0.10	0.06	11435	0.67	92	11626	0.66	36	0.02
ITEM 24	0.94	3.97	0.05	0.14	A	0.07	0.12	0.06	11507	0.61	94	11663	0.61	38	0.00
ITEM 25	1.92	386.33	0.00	-1.53	C	0.08	-1.11	0.06	11556	0.72	98	11702	0.61	40	0.11
ITEM 26	1.09	8.08	0.00	-0.21	A	0.07	-0.17	0.06	11440	0.66	92	11546	0.64	37	0.03
ITEM 27	1.19	25.11	0.00	-0.41	A	0.08	-0.26	0.06	11446	0.65	92	11585	0.61	37	0.04
ITEM 28	0.77	75.61	0.00	0.62	A	0.07	0.48	0.06	11515	0.56	92	11714	0.60	37	-0.04
ITEM 29	0.81	42.73	0.00	0.48	A	0.07	0.35	0.06	11429	0.53	93	11586	0.56	37	-0.02
ITEM 30	0.81	52.14	0.00	0.50	A	0.07	0.42	0.06	11213	0.53	99	11433	0.56	43	-0.03
ITEM 31	1.18	24.60	0.00	-0.38	A	0.08	-0.27	0.06	11301	0.64	103	11520	0.60	48	0.04
ITEM 32	0.94	4.25	0.04	0.15	A	0.07	0.13	0.06	11317	0.69	91	11480	0.69	36	0.00
ITEM 33	1.09	9.47	0.00	-0.20	A	0.06	-0.19	0.06	11462	0.49	91	11630	0.46	36	0.03
ITEM 34	0.84	30.59	0.00	0.40	A	0.07	0.30	0.06	11188	0.54	94	11297	0.56	38	-0.02
ITEM 35	1.25	59.71	0.00	-0.53	A	0.07	-0.45	0.06	11280	0.62	91	11472	0.56	36	0.06
ITEM 36	1.22	39.43	0.00	-0.47	A	0.07	-0.34	0.06	11347	0.45	95	11539	0.40	41	0.05
ITEM 37	1.15	18.62	0.00	-0.33	A	0.08	-0.23	0.06	11402	0.48	94	11592	0.44	42	0.04
ITEM 38	1.17	31.65	0.00	-0.37	A	0.07	-0.34	0.06	11273	0.46	95	11398	0.41	39	0.05
ITEM 39	0.88	13.76	0.00	0.29	A	0.08	0.19	0.06	11318	0.41	94	11490	0.41	40	0.00
ITEM 40	0.93	4.82	0.03	0.18	A	0.08	0.13	0.06	11125	0.31	101	11227	0.31	43	0.01

Source: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988: Base Year Survey .

Appendix B-3

Differential Item Functioning (DIF), Science

MANTEL-HAENSZEL ODDS-RATIO AND OTHER STATISTICS, NUMBER OF TABLES = 25

		NO. LEVELS	LEVEL 1		LEVEL 2	
GROUP VARIABLE:	RACE	2	WHITE	(REFERENCE)	ASIAN	(FOCAL)
RESPONSE VARIABLE:	ITEMSCOR	2	RIGHT		WRONG	
STRATIFYING VARIABLE:	# RIGHT	26				

	MH ODDS RATIO	MH CHI-SQUARE	PROB > CHI-SQ	MH D-DIF		STD ERR MH D-DIF	STDZD D-DIF	STD ERR STD D-DIF	REFERENCE N	REFERENCE P+	REFERENCE N0*	FOCAL N	FOCAL P+	FOCAL N0*	IMPACT
ITEM 1	1.36	21.38	0.00	-0.73	A	0.16	-0.59	0.14	15708	0.75	31	1488	0.70	4	0.06
ITEM 2	0.91	1.25	0.26	0.21	A	0.18	0.18	0.17	15698	0.82	31	1488	0.83	4	-0.01
ITEM 3	0.95	0.58	0.45	0.12	A	0.15	0.11	0.14	15630	0.69	31	1477	0.70	4	-0.01
ITEM 4	1.18	6.79	0.01	-0.38	A	0.15	-0.34	0.14	15677	0.71	40	1481	0.67	7	0.03
ITEM 5	1.02	0.04	0.84	-0.05	A	0.19	-0.03	0.16	15673	0.81	386	1479	0.80	54	0.01
ITEM 6	1.32	12.56	0.00	-0.66	A	0.18	-0.47	0.15	15649	0.82	386	1478	0.77	54	0.04
ITEM 7	0.82	9.41	0.00	0.47	A	0.15	0.41	0.14	15636	0.69	136	1481	0.72	19	-0.03
ITEM 8	1.47	31.34	0.00	-0.91	A	0.14	-0.77	0.13	15707	0.61	31	1483	0.53	4	0.08
ITEM 9	0.99	0.03	0.87	0.03	A	0.15	0.02	0.14	15693	0.68	31	1487	0.68	4	0.00
ITEM 10	1.04	0.39	0.53	-0.10	A	0.15	-0.08	0.13	15513	0.60	41	1472	0.59	7	0.01
ITEM 11	0.85	7.22	0.01	0.37	A	0.14	0.33	0.13	15447	0.52	31	1464	0.55	4	-0.04
ITEM 12	0.95	0.46	0.50	0.12	A	0.17	0.09	0.15	14885	0.75	136	1422	0.75	19	0.00
ITEM 13	0.83	5.84	0.02	0.43	A	0.18	0.36	0.16	15397	0.78	30	1455	0.80	4	-0.02
ITEM 14	1.25	11.26	0.00	-0.53	A	0.16	-0.37	0.13	15692	0.63	136	1484	0.59	19	0.04
ITEM 15	0.82	11.85	0.00	0.48	A	0.14	0.41	0.13	15552	0.43	39	1465	0.47	7	-0.04
ITEM 16	0.85	7.70	0.01	0.39	A	0.14	0.33	0.13	15510	0.50	31	1460	0.54	4	-0.04
ITEM 17	0.97	0.29	0.59	0.08	A	0.14	0.06	0.13	15582	0.47	136	1464	0.48	19	-0.01
ITEM 18	1.33	22.12	0.00	-0.68	A	0.15	-0.54	0.13	15528	0.52	31	1459	0.46	4	0.06
ITEM 19	0.88	4.46	0.03	0.31	A	0.14	0.25	0.13	15581	0.47	31	1472	0.50	4	-0.03
ITEM 20	0.94	0.89	0.35	0.13	A	0.14	0.12	0.13	15545	0.45	39	1460	0.46	7	-0.02
ITEM 21	0.93	1.54	0.22	0.17	A	0.14	0.15	0.13	15537	0.46	39	1463	0.48	7	-0.02
ITEM 22	0.74	26.56	0.00	0.70	A	0.14	0.63	0.13	15443	0.40	31	1440	0.47	4	-0.07
ITEM 23	0.97	0.19	0.67	0.06	A	0.14	0.06	0.13	15182	0.43	31	1420	0.44	4	-0.01
ITEM 24	1.12	2.91	0.09	-0.26	A	0.15	-0.21	0.13	15530	0.38	40	1452	0.36	7	0.02
ITEM 25	1.05	0.45	0.50	-0.11	A	0.16	-0.10	0.15	15470	0.24	31	1448	0.24	4	0.00

Source: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988: Base Year Survey .

Appendix B-3--(continued)

Differential Item Functioning (DIF), Science

MANTEL-HAENSZEL ODDS-RATIO AND OTHER STATISTICS, NUMBER OF TABLES = 25

		NO. LEVELS	LEVEL 1		LEVEL 2	
GROUP VARIABLE:	RACE	2	WHITE	(REFERENCE)	HISPANIC	(FOCAL)
RESPONSE VARIABLE:	ITEMSCOR	2	RIGHT		WRONG	
STRATIFYING VARIABLE:	# RIGHT	26				

	MH ODDS RATIO	MH CHI-SQUARE	PROB > CHI-SQ	MH D-DIF		STD ERR MH D-DIF	STDZD D-DIF	STD ERR STD D-DIF	REFERENCE N	REFERENCE P+	REFERENCE NO#	FOCAL N	FOCAL P+	FOCAL NO#	IMPACT
ITEM 1	0.96	0.76	0.38	0.10	A	0.11	0.08	0.10	15708	0.75	31	2975	0.64	9	0.11
ITEM 2	1.02	0.10	0.76	-0.04	A	0.12	0.05	0.11	15698	0.82	26	2979	0.73	3	0.09
ITEM 3	0.99	0.09	0.76	0.04	A	0.11	0.03	0.10	15630	0.69	31	2955	0.58	9	0.11
ITEM 4	0.92	3.60	0.06	0.20	A	0.11	0.20	0.10	15677	0.71	35	2957	0.63	9	0.08
ITEM 5	1.04	0.47	0.49	-0.09	A	0.12	-0.07	0.10	15673	0.81	386	2959	0.67	22	0.14
ITEM 6	1.22	15.00	0.00	-0.47	A	0.12	-0.35	0.10	15649	0.82	386	2943	0.66	22	0.16
ITEM 7	0.79	26.38	0.00	0.57	A	0.11	0.47	0.10	15636	0.69	136	2951	0.62	11	0.07
ITEM 8	1.05	1.15	0.28	-0.11	A	0.10	-0.12	0.10	15707	0.61	26	2972	0.49	3	0.13
ITEM 9	0.97	0.33	0.57	0.06	A	0.11	0.05	0.10	15693	0.68	41	2977	0.57	15	0.12
ITEM 10	1.24	23.08	0.00	-0.51	A	0.11	-0.43	0.10	15513	0.60	36	2939	0.42	9	0.18
ITEM 11	0.92	3.25	0.07	0.19	A	0.10	0.15	0.10	15447	0.52	26	2910	0.43	3	0.09
ITEM 12	1.06	1.56	0.21	-0.15	A	0.11	-0.12	0.10	14885	0.75	136	2832	0.60	10	0.15
ITEM 13	0.91	3.47	0.06	0.22	A	0.12	0.18	0.11	15397	0.78	26	2892	0.69	3	0.09
ITEM 14	1.65	114.67	0.00	-1.17	B	0.11	-0.92	0.10	15692	0.63	136	2962	0.37	10	0.26
ITEM 15	0.76	36.97	0.00	0.66	A	0.11	0.58	0.10	15552	0.43	31	2909	0.37	8	0.06
ITEM 16	0.86	11.03	0.00	0.34	A	0.10	0.33	0.10	15510	0.50	31	2921	0.43	8	0.07
ITEM 17	1.02	0.26	0.61	-0.06	A	0.11	-0.04	0.10	15582	0.47	136	2925	0.34	10	0.13
ITEM 18	1.15	9.42	0.00	-0.33	A	0.11	-0.26	0.10	15528	0.52	26	2913	0.35	3	0.16
ITEM 19	1.10	4.06	0.04	-0.22	A	0.11	-0.19	0.10	15581	0.47	34	2924	0.33	9	0.14
ITEM 20	0.94	1.87	0.17	0.14	A	0.10	0.14	0.10	15545	0.45	31	2907	0.37	8	0.07
ITEM 21	0.95	1.35	0.24	0.12	A	0.10	0.12	0.10	15537	0.46	34	2913	0.37	8	0.09
ITEM 22	0.87	9.30	0.00	0.33	A	0.11	0.30	0.10	15443	0.40	31	2890	0.34	8	0.06
ITEM 23	1.05	1.28	0.26	-0.12	A	0.10	-0.11	0.10	15182	0.43	31	2853	0.35	8	0.08
ITEM 24	0.97	0.34	0.56	0.07	A	0.12	0.05	0.11	15530	0.38	40	2900	0.25	14	0.12
ITEM 25	0.99	0.04	0.84	0.03	A	0.13	0.03	0.13	15470	0.24	26	2878	0.18	3	0.07

Source: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988: Base Year Survey .

Appendix B-3--(continued)

Differential Item Functioning (DIF), Science

MANTEL-HAENSZEL ODDS-RATIO AND OTHER STATISTICS, NUMBER OF TABLES = 25

		NO. LEVELS	LEVEL 1		LEVEL 2	
GROUP VARIABLE:	RACE	2	WHITE	(REFERENCE)	BLACK	(FOCAL)
RESPONSE VARIABLE:	ITEMSCOR	2	RIGHT		WRONG	
STRATIFYING VARIABLE:	# RIGHT	26				

	MH ODDS RATIO	MH CHI-SQUARE	PROB > CHI-SQ	MH D-DIF		STD ERR MH D-DIF	STDZD D-DIF	STD ERR STD D-DIF	REFERENCE N	REFERENCE P+	REFERENCE NO*	FOCAL N	FOCAL P+	FOCAL NO*	IMPACT
ITEM 1	1.45	63.91	0.00	-0.88	A	0.11	-0.76	0.10	15708	0.75	31	2828	0.51	1	0.24
ITEM 2	1.04	0.53	0.47	-0.09	A	0.12	-0.08	0.11	15698	0.82	26	2830	0.70	0	0.13
ITEM 3	1.01	0.01	0.91	-0.01	A	0.11	0.01	0.10	15630	0.69	31	2790	0.54	1	0.15
ITEM 4	1.00	0.01	0.94	0.01	A	0.11	0.01	0.10	15677	0.71	31	2822	0.57	2	0.13
ITEM 5	1.16	7.53	0.01	-0.34	A	0.12	-0.25	0.10	15673	0.81	136	2820	0.60	5	0.21
ITEM 6	0.87	6.38	0.01	0.32	A	0.12	0.26	0.11	15649	0.82	386	2815	0.66	7	0.15
ITEM 7	0.89	6.25	0.01	0.28	A	0.11	0.24	0.10	15636	0.69	136	2811	0.56	4	0.13
ITEM 8	0.89	6.10	0.01	0.26	A	0.11	0.25	0.10	15707	0.61	26	2822	0.49	0	0.12
ITEM 9	0.88	7.55	0.01	0.31	A	0.11	0.27	0.10	15693	0.68	41	2817	0.55	8	0.13
ITEM 10	0.93	2.06	0.15	0.16	A	0.11	0.15	0.10	15513	0.60	41	2781	0.44	8	0.16
ITEM 11	0.83	15.52	0.00	0.43	A	0.11	0.39	0.10	15447	0.52	31	2749	0.43	1	0.09
ITEM 12	1.05	0.97	0.32	-0.12	A	0.12	-0.08	0.10	14885	0.75	136	2699	0.56	4	0.19
ITEM 13	0.88	5.65	0.02	0.29	A	0.12	0.24	0.11	15397	0.78	26	2691	0.66	0	0.12
ITEM 14	2.30	271.47	0.00	-1.96	C	0.12	-1.59	0.11	15692	0.63	136	2814	0.27	5	0.36
ITEM 15	0.93	2.31	0.13	0.18	A	0.12	0.16	0.11	15552	0.43	31	2741	0.30	1	0.13
ITEM 16	0.82	18.96	0.00	0.47	A	0.11	0.43	0.10	15510	0.50	26	2753	0.41	0	0.09
ITEM 17	0.96	0.67	0.41	0.10	A	0.11	0.12	0.11	15582	0.47	31	2759	0.32	1	0.15
ITEM 18	1.18	11.15	0.00	-0.39	A	0.12	-0.31	0.11	15528	0.52	31	2741	0.31	2	0.21
ITEM 19	0.94	1.70	0.19	0.15	A	0.11	0.10	0.11	15581	0.47	39	2750	0.33	8	0.14
ITEM 20	0.85	11.43	0.00	0.37	A	0.11	0.34	0.11	15545	0.45	31	2722	0.37	2	0.08
ITEM 21	0.83	15.86	0.00	0.43	A	0.11	0.41	0.11	15537	0.46	26	2719	0.37	0	0.08
ITEM 22	0.89	6.04	0.01	0.28	A	0.11	0.24	0.11	15443	0.40	31	2695	0.31	2	0.09
ITEM 23	0.95	1.12	0.29	0.12	A	0.11	0.10	0.11	15182	0.43	31	2651	0.35	2	0.07
ITEM 24	1.04	0.37	0.54	-0.08	A	0.13	-0.08	0.12	15530	0.38	40	2684	0.21	9	0.17
ITEM 25	1.00	0.00	1.00	0.00	A	0.14	-0.01	0.14	15470	0.24	26	2678	0.16	0	0.09

Source: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988: Base Year Survey .

Differential Item Functioning (DIF), Science

MANTEL-HAENSZEL ODDS-RATIO AND OTHER STATISTICS, NUMBER OF TABLES = 25

		NO. LEVELS	LEVEL 1		LEVEL 2	
GROUP VARIABLE:	RACE	2	WHITE	(REFERENCE)	AM IND	(FOCAL)
RESPONSE VARIABLE:	ITEMSCOR	2	RIGHT		WRONG	
STRATIFYING VARIABLE:	# RIGHT	26				

	MH ODDS RATIO	MH CHI-SQUARE	PROB > CHI-SQ	MH D-DIF		STD ERR MH D-DIF	STDZD D-DIF	STD ERR STD D-DIF	REFERENCE N	REFERENCE P+	REFERENCE NO#	FOCAL N	FOCAL P+	FOCAL NO#	IMPACT
ITEM 1	1.20	1.86	0.17	-0.43	A	0.30	-0.36	0.28	15708	0.75	386	305	0.55	1	0.21
ITEM 2	1.08	0.21	0.65	-0.18	A	0.33	-0.15	0.29	15698	0.82	381	302	0.67	0	0.16
ITEM 3	0.91	0.52	0.47	0.23	A	0.30	0.19	0.28	15630	0.69	386	301	0.55	1	0.14
ITEM 4	1.06	0.14	0.71	-0.13	A	0.30	-0.12	0.28	15677	0.71	395	301	0.55	5	0.16
ITEM 5	0.87	0.72	0.40	0.33	A	0.35	0.22	0.29	15673	0.81	386	301	0.63	1	0.18
ITEM 6	1.03	0.02	0.90	-0.07	A	0.34	-0.05	0.28	15649	0.82	386	298	0.62	1	0.20
ITEM 7	0.85	1.43	0.23	0.40	A	0.31	0.30	0.28	15636	0.69	385	300	0.56	1	0.13
ITEM 8	0.99	0.00	0.98	0.03	A	0.30	0.03	0.28	15707	0.61	386	304	0.46	1	0.15
ITEM 9	1.06	0.15	0.70	-0.13	A	0.30	-0.11	0.28	15693	0.68	386	301	0.50	1	0.18
ITEM 10	1.09	0.33	0.57	-0.20	A	0.31	-0.16	0.28	15513	0.60	396	299	0.41	3	0.19
ITEM 11	1.01	0.00	0.99	-0.02	A	0.30	-0.01	0.29	15447	0.52	395	294	0.38	3	0.14
ITEM 12	0.80	2.47	0.12	0.52	A	0.32	0.45	0.29	14885	0.75	1272	285	0.60	1	0.14
ITEM 13	0.90	0.48	0.49	0.26	A	0.34	0.19	0.29	15397	0.78	380	288	0.64	1	0.14
ITEM 14	1.54	9.79	0.00	-1.02	B	0.33	-0.82	0.29	15692	0.63	386	297	0.34	1	0.29
ITEM 15	0.94	0.15	0.70	0.15	A	0.33	0.13	0.30	15552	0.43	394	294	0.30	2	0.13
ITEM 16	0.99	0.00	0.98	0.01	A	0.30	0.02	0.29	15510	0.50	386	298	0.36	1	0.14
ITEM 17	0.90	0.54	0.46	0.25	A	0.31	0.23	0.29	15582	0.47	386	300	0.33	1	0.13
ITEM 18	1.00	0.00	0.96	0.00	A	0.32	-0.01	0.29	15528	0.52	386	298	0.34	2	0.18
ITEM 19	1.07	0.19	0.67	-0.16	A	0.32	-0.15	0.30	15581	0.47	394	300	0.30	4	0.17
ITEM 20	1.01	0.00	0.99	-0.02	A	0.30	0.00	0.29	15545	0.45	394	297	0.33	2	0.11
ITEM 21	0.88	0.85	0.36	0.29	A	0.30	0.29	0.29	15537	0.46	389	296	0.36	2	0.10
ITEM 22	1.13	0.68	0.41	-0.30	A	0.33	-0.29	0.32	15443	0.40	385	295	0.26	1	0.14
ITEM 23	0.77	4.27	0.04	0.62	A	0.29	0.60	0.29	15182	0.43	376	293	0.40	1	0.03
ITEM 24	1.17	0.87	0.35	-0.37	A	0.37	-0.32	0.35	15530	0.38	395	296	0.20	3	0.18
ITEM 25	0.97	0.01	0.94	0.06	A	0.39	0.05	0.38	15470	0.24	379	294	0.16	1	0.08

Source: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988: Base Year Survey .

Appendix B-3--(continued)

Differential Item Functioning (DIF), Science

MANTEL-HAENSZEL ODDS-RATIO AND OTHER STATISTICS, NUMBER OF TABLES = 25

		NO. LEVELS	LEVEL 1		LEVEL 2	
GROUP VARIABLE:	SEX	2	MALE	(REFERENCE)	FEMALE	(FOCAL)
RESPONSE VARIABLE:	ITEMSCOR	2	RIGHT		WRONG	
STRATIFYING VARIABLE:	# RIGHT	26				

	MH ODDS RATIO	MH CHI-SQUARE	PROB > CHI-SQ	MH D-DIF		STD ERR MH D-DIF	STDZD D-DIF	STD ERR STD D-DIF	REFERENCE N	REFERENCE P+	REFERENCE NO*	FOCAL N	FOCAL P+	FOCAL NO*	IMPACT
ITEM 1	0.86	20.47	0.00	0.34	A	0.08	0.29	0.07	11617	0.70	34	11737	0.70	12	0.00
ITEM 2	1.29	53.34	0.00	-0.61	A	0.08	-0.51	0.07	11610	0.82	25	11739	0.77	8	0.05
ITEM 3	0.97	0.82	0.37	0.07	A	0.07	0.05	0.06	11538	0.66	34	11666	0.65	12	0.01
ITEM 4	0.67	173.63	0.00	0.93	A	0.07	0.84	0.07	11580	0.65	25	11709	0.70	8	-0.06
ITEM 5	1.19	21.28	0.00	-0.41	A	0.09	-0.29	0.07	11583	0.78	124	11699	0.75	47	0.04
ITEM 6	0.94	3.08	0.08	0.15	A	0.09	0.11	0.07	11550	0.77	332	11682	0.77	137	0.01
ITEM 7	1.58	227.68	0.00	-1.08	B	0.07	-0.91	0.06	11553	0.72	124	11677	0.61	46	0.10
ITEM 8	1.27	71.68	0.00	-0.57	A	0.07	-0.49	0.06	11628	0.61	25	11714	0.54	8	0.07
ITEM 9	0.89	14.29	0.00	0.27	A	0.07	0.21	0.06	11609	0.65	34	11715	0.65	12	0.00
ITEM 10	1.00	0.00	1.00	0.00	A	0.07	0.00	0.06	11441	0.57	44	11610	0.54	20	0.03
ITEM 11	1.14	21.98	0.00	-0.31	A	0.07	-0.26	0.06	11370	0.52	25	11544	0.47	8	0.05
ITEM 12	1.56	173.12	0.00	-1.05	B	0.08	-0.84	0.07	10997	0.75	123	11175	0.66	46	0.09
ITEM 13	0.77	60.73	0.00	0.62	A	0.08	0.52	0.07	11209	0.74	25	11563	0.77	9	-0.03
ITEM 14	1.39	109.78	0.00	-0.77	A	0.07	-0.57	0.06	11589	0.60	123	11706	0.50	47	0.10
ITEM 15	0.73	118.38	0.00	0.75	A	0.07	0.66	0.06	11431	0.39	33	11573	0.43	12	-0.04
ITEM 16	0.87	23.97	0.00	0.33	A	0.07	0.29	0.06	11448	0.48	25	11538	0.48	8	0.00
ITEM 17	1.12	14.60	0.00	-0.26	A	0.07	-0.21	0.06	11488	0.46	33	11583	0.40	12	0.06
ITEM 18	1.29	72.48	0.00	-0.59	A	0.07	-0.47	0.06	11429	0.51	25	11565	0.42	9	0.09
ITEM 19	0.95	3.04	0.08	0.12	A	0.07	0.09	0.06	11448	0.45	25	11625	0.42	9	0.02
ITEM 20	0.93	6.26	0.01	0.17	A	0.07	0.14	0.06	11413	0.43	33	11564	0.42	13	0.01
ITEM 21	1.10	10.32	0.00	-0.22	A	0.07	-0.18	0.06	11406	0.46	25	11572	0.41	9	0.05
ITEM 22	0.74	107.52	0.00	0.71	A	0.07	0.65	0.06	11365	0.36	33	11449	0.40	13	-0.04
ITEM 23	0.97	0.99	0.32	0.07	A	0.07	0.06	0.06	11218	0.42	25	11232	0.40	9	0.01
ITEM 24	0.93	5.03	0.02	0.17	A	0.07	0.13	0.06	11401	0.35	33	11504	0.33	13	0.02
ITEM 25	0.79	48.97	0.00	0.55	A	0.08	0.51	0.08	11329	0.22	25	11486	0.23	9	-0.01

Source: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988: Base Year Survey .

Appendix B-4

Differential Item Functioning (DIF), History/Citizenship/Geography

MANTEL-HAENSZEL ODDS-RATIO AND OTHER STATISTICS, NUMBER OF TABLES = 30

		NO. LEVELS	LEVEL 1		LEVEL 2	
GROUP VARIABLE:	RACE	2	WHITE	(REFERENCE)	ASIAN	(FOCAL)
RESPONSE VARIABLE:	ITEMSCOR	2	RIGHT		WRONG	
STRATIFYING VARIABLE:	# RIGHT	31				

	MH ODDS RATIO	MH CHI-SQUARE	PROB > CHI-SQ	MH D-DIF		STD ERR MH D-DIF	STDZD D-DIF	STD ERR STD D-DIF	REFERENCE N	REFERENCE P+	REFERENCE N0#	FOCAL N	FOCAL P+	FOCAL N0#	IMPACT
ITEM 1	0.87	2.42	0.12	0.33	A	0.21	0.28	0.19	15457	0.85	208	1463	0.87	33	-0.02
ITEM 2	1.52	28.58	0.00	-0.98	A	0.18	-0.70	0.15	15668	0.82	208	1483	0.77	33	0.04
ITEM 3	1.24	3.32	0.07	-0.50	A	0.28	-0.37	0.23	15677	0.93	2114	1480	0.92	242	0.01
ITEM 4	2.00	109.97	0.00	-1.63	C	0.16	-1.23	0.13	15628	0.76	208	1477	0.66	33	0.10
ITEM 5	1.21	3.68	0.06	-0.45	A	0.23	-0.37	0.20	15581	0.90	633	1474	0.89	83	0.01
ITEM 6	1.14	2.10	0.15	-0.30	A	0.20	-0.26	0.19	15595	0.87	218	1471	0.86	33	0.01
ITEM 7	2.16	37.24	0.00	-1.81	C	0.30	-1.12	0.23	15594	0.95	1966	1470	0.91	216	0.04
ITEM 8	1.49	14.51	0.00	-0.94	A	0.24	-0.70	0.21	15583	0.92	837	1468	0.89	99	0.03
ITEM 9	3.10	89.63	0.00	-2.66	C	0.29	-1.67	0.22	15596	0.95	2206	1471	0.90	240	0.05
ITEM 10	1.01	0.01	0.93	-0.02	A	0.16	-0.01	0.14	15638	0.73	208	1477	0.74	33	-0.01
ITEM 11	0.98	0.06	0.81	0.04	A	0.16	0.03	0.13	15637	0.65	208	1474	0.67	33	-0.02
ITEM 12	0.66	43.15	0.00	0.97	A	0.15	0.81	0.14	15623	0.59	208	1470	0.68	33	-0.09
ITEM 13	1.14	4.19	0.04	-0.32	A	0.15	-0.24	0.13	15560	0.63	208	1465	0.62	33	0.01
ITEM 14	0.47	156.24	0.00	1.77	C	0.14	1.48	0.13	15541	0.44	240	1471	0.61	42	-0.17
ITEM 15	0.95	0.64	0.42	0.12	A	0.15	0.09	0.13	15654	0.52	208	1483	0.55	33	-0.03
ITEM 16	0.76	19.64	0.00	0.64	A	0.14	0.52	0.13	15643	0.48	208	1481	0.56	33	-0.07
ITEM 17	1.60	33.42	0.00	-1.11	B	0.20	-0.86	0.17	15634	0.87	208	1473	0.82	33	0.05
ITEM 18	1.05	0.32	0.57	-0.11	A	0.19	-0.08	0.16	15653	0.82	208	1480	0.82	33	0.00
ITEM 19	0.60	29.21	0.00	1.21	B	0.23	0.79	0.18	15630	0.81	623	1475	0.86	83	-0.05
ITEM 20	1.43	29.13	0.00	-0.84	A	0.16	-0.65	0.14	15609	0.72	208	1480	0.67	33	0.05
ITEM 21	0.65	33.18	0.00	1.01	B	0.18	0.79	0.16	15590	0.72	208	1474	0.79	33	-0.07
ITEM 22	0.73	25.68	0.00	0.74	A	0.15	0.57	0.13	15581	0.53	208	1475	0.61	33	-0.08
ITEM 23	0.96	0.35	0.55	0.09	A	0.14	0.07	0.13	15593	0.51	208	1469	0.54	33	-0.03
ITEM 24	1.10	2.43	0.12	-0.23	A	0.14	-0.18	0.13	15557	0.58	208	1472	0.57	33	0.00
ITEM 25	0.95	0.61	0.44	0.12	A	0.14	0.09	0.13	15376	0.52	220	1452	0.55	33	-0.03
ITEM 26	1.24	12.79	0.00	-0.51	A	0.14	-0.41	0.13	15559	0.55	221	1467	0.53	33	0.02
ITEM 27	0.93	1.23	0.27	0.17	A	0.15	0.13	0.13	15517	0.57	221	1460	0.61	33	-0.03
ITEM 28	1.12	3.02	0.08	-0.26	A	0.15	-0.20	0.13	15496	0.48	221	1450	0.48	33	0.00
ITEM 29	0.91	2.57	0.11	0.23	A	0.14	0.19	0.13	15530	0.38	221	1459	0.42	33	-0.04
ITEM 30	0.91	1.73	0.19	0.21	A	0.16	0.18	0.14	15472	0.26	221	1454	0.29	33	-0.03

Source: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988: Base Year Survey .

Appendix B-4--(continued)

Differential Item Functioning (DIF), History/Citizenship/Geography

MANTEL-HAENSZEL ODDS-RATIO AND OTHER STATISTICS, NUMBER OF TABLES = 30

		NO. LEVELS	LEVEL 1	LEVEL 2
GROUP VARIABLE:	RACE	2	WHITE (REFERENCE)	HISPANIC (FOCAL)
RESPONSE VARIABLE:	ITEMSCOR	2	RIGHT	WRONG
STRATIFYING VARIABLE:	# RIGHT	31		

	MH ODDS RATIO	MH CHI-SQUARE	PROB > CHI-SQ	MH D-DIF		STD ERR MH D-DIF	STDZD D-DIF	STD ERR STD D-DIF	REFERENCE N	REFERENCE P+	REFERENCE ND*	FOCAL N	FOCAL P+	FOCAL ND*	IMPACT
ITEM 1	1.00	0.00	0.98	0.00	A	0.13	0.00	0.12	15457	0.85	205	2920	0.76	12	0.09
ITEM 2	1.35	35.75	0.00	-0.71	A	0.12	-0.53	0.10	15668	0.82	205	2966	0.65	12	0.17
ITEM 3	1.08	1.11	0.29	-0.18	A	0.16	-0.10	0.14	15677	0.93	2111	2969	0.85	145	0.08
ITEM 4	1.79	158.40	0.00	-1.38	B	0.11	-1.05	0.10	15628	0.76	208	2955	0.52	14	0.23
ITEM 5	1.10	2.55	0.11	-0.23	A	0.14	-0.18	0.13	15581	0.90	620	2931	0.82	29	0.09
ITEM 6	1.29	21.30	0.00	-0.60	A	0.13	-0.51	0.12	15595	0.87	205	2933	0.78	11	0.10
ITEM 7	1.76	59.58	0.00	-1.34	B	0.18	-0.90	0.14	15594	0.95	1955	2930	0.84	117	0.11
ITEM 8	1.40	28.70	0.00	-0.80	A	0.15	-0.62	0.13	15583	0.92	837	2934	0.81	53	0.11
ITEM 9	2.27	127.05	0.00	-1.93	C	0.18	-1.31	0.14	15596	0.95	1252	2930	0.83	71	0.12
ITEM 10	0.79	24.34	0.00	0.55	A	0.11	0.49	0.10	15638	0.73	205	2958	0.67	12	0.06
ITEM 11	1.08	2.60	0.11	-0.18	A	0.11	-0.12	0.10	15637	0.65	208	2951	0.49	14	0.16
ITEM 12	0.88	7.31	0.01	0.29	A	0.11	0.24	0.10	15623	0.59	208	2949	0.49	15	0.10
ITEM 13	0.83	17.14	0.00	0.45	A	0.11	0.37	0.10	15560	0.63	208	2926	0.53	15	0.10
ITEM 14	0.43	366.34	0.00	1.98	C	0.11	1.74	0.10	15541	0.44	205	2935	0.52	12	-0.08
ITEM 15	0.89	6.54	0.01	0.28	A	0.11	0.24	0.10	15654	0.52	205	2957	0.41	11	0.12
ITEM 16	0.88	7.57	0.01	0.29	A	0.11	0.24	0.10	15643	0.48	208	2950	0.40	14	0.08
ITEM 17	1.25	16.25	0.00	-0.53	A	0.13	-0.42	0.11	15634	0.87	205	2951	0.74	11	0.13
ITEM 18	1.07	1.89	0.17	-0.17	A	0.12	-0.13	0.11	15653	0.82	205	2946	0.71	11	0.11
ITEM 19	0.83	11.68	0.00	0.44	A	0.13	0.34	0.10	15630	0.81	623	2948	0.70	32	0.11
ITEM 20	1.18	12.96	0.00	-0.39	A	0.11	-0.30	0.10	15609	0.72	208	2940	0.56	13	0.16
ITEM 21	1.02	0.11	0.74	-0.04	A	0.11	-0.02	0.10	15590	0.72	205	2938	0.59	11	0.13
ITEM 22	0.79	26.24	0.00	0.55	A	0.11	0.47	0.10	15581	0.53	205	2924	0.44	11	0.09
ITEM 23	0.85	13.75	0.00	0.39	A	0.11	0.34	0.10	15593	0.51	205	2939	0.43	11	0.08
ITEM 24	0.90	4.79	0.03	0.24	A	0.11	0.20	0.10	15557	0.58	208	2929	0.47	13	0.11
ITEM 25	0.96	0.71	0.40	0.09	A	0.11	0.09	0.10	15376	0.52	220	2884	0.41	20	0.11
ITEM 26	1.26	26.83	0.00	-0.55	A	0.11	-0.46	0.10	15559	0.55	205	2919	0.38	11	0.17
ITEM 27	1.04	0.53	0.47	-0.08	A	0.11	-0.05	0.10	15517	0.57	208	2906	0.42	14	0.15
ITEM 28	1.04	0.83	0.36	-0.10	A	0.11	-0.08	0.10	15496	0.48	208	2892	0.36	13	0.11
ITEM 29	0.96	0.77	0.38	0.10	A	0.11	0.08	0.11	15530	0.38	208	2897	0.31	12	0.07
ITEM 30	0.86	8.82	0.00	0.36	A	0.12	0.34	0.12	15472	0.26	208	2888	0.23	12	0.03

Source: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988: Base Year Survey .

Appendix B-4--(continued)

Differential Item Functioning (DIF), History/Citizenship/Geography

MANTEL-HAENSZEL ODDS-RATIO AND OTHER STATISTICS, NUMBER OF TABLES = 30

		NO. LEVELS	LEVEL 1		LEVEL 2	
GROUP VARIABLE:	RACE	2	WHITE	(REFERENCE)	BLACK	(FOCAL)
RESPONSE VARIABLE:	ITEMSCOR	2	RIGHT		WRONG	
STRATIFYING VARIABLE:	# RIGHT	31				

	MH ODDS RATIO	MH CHI-SQUARE	PROB > CHI-SQ	MH D-DIF	STD ERR MH D-DIF	STDZD D-DIF	STD ERR STD D-DIF	REFERENCE N	REFERENCE P+	REFERENCE NO*	FOCAL N	FOCAL P+	FOCAL NO*	IMPACT
ITEM 1	1.40	43.09	0.00	-0.80 A	0.12	-0.67	0.11	15457	0.85	205	2763	0.69	8	0.16
ITEM 2	0.65	58.40	0.00	1.01 B	0.13	0.79	0.11	15668	0.82	205	2839	0.75	8	0.07
ITEM 3	1.17	5.02	0.03	-0.37 A	0.16	-0.27	0.14	15677	0.93	2111	2838	0.83	94	0.10
ITEM 4	1.32	31.90	0.00	-0.65 A	0.11	-0.52	0.10	15628	0.76	205	2810	0.56	8	0.20
ITEM 5	1.08	1.78	0.18	-0.19 A	0.14	-0.15	0.13	15581	0.90	620	2790	0.81	18	0.09
ITEM 6	1.06	0.98	0.32	-0.14 A	0.14	-0.13	0.13	15595	0.87	208	2800	0.80	7	0.08
ITEM 7	1.54	31.94	0.00	-1.02 B	0.18	-0.72	0.15	15594	0.95	1955	2792	0.85	69	0.09
ITEM 8	0.99	0.02	0.90	0.03 A	0.16	0.02	0.14	15583	0.92	837	2798	0.85	28	0.07
ITEM 9	1.60	36.29	0.00	-1.11 B	0.19	-0.79	0.15	15596	0.95	2196	2798	0.86	108	0.09
ITEM 10	0.89	5.60	0.02	0.26 A	0.11	0.24	0.10	15638	0.73	205	2827	0.63	8	0.10
ITEM 11	0.99	0.01	0.92	0.01 A	0.11	0.04	0.10	15637	0.65	208	2820	0.48	9	0.18
ITEM 12	0.88	7.67	0.01	0.30 A	0.11	0.26	0.10	15623	0.59	205	2811	0.47	8	0.12
ITEM 13	0.74	39.67	0.00	0.69 A	0.11	0.58	0.10	15560	0.63	208	2795	0.52	10	0.10
ITEM 14	0.79	26.64	0.00	0.57 A	0.11	0.50	0.10	15541	0.44	217	2797	0.37	14	0.07
ITEM 15	1.06	1.49	0.22	-0.14 A	0.11	-0.11	0.10	15654	0.52	208	2822	0.34	10	0.18
ITEM 16	0.97	0.52	0.47	0.08 A	0.11	0.08	0.10	15643	0.48	205	2816	0.36	8	0.12
ITEM 17	0.54	87.85	0.00	1.45 B	0.15	1.13	0.13	15634	0.87	208	2822	0.84	10	0.03
ITEM 18	1.07	1.78	0.18	-0.16 A	0.12	-0.13	0.11	15653	0.82	208	2825	0.69	9	0.13
ITEM 19	0.98	0.12	0.73	0.05 A	0.13	0.03	0.10	15630	0.81	623	2815	0.66	21	0.16
ITEM 20	1.23	19.13	0.00	-0.48 A	0.11	-0.38	0.10	15609	0.72	208	2801	0.53	9	0.19
ITEM 21	1.38	48.13	0.00	-0.76 A	0.11	-0.63	0.10	15590	0.72	205	2799	0.51	7	0.21
ITEM 22	1.06	1.60	0.21	-0.15 A	0.11	-0.12	0.10	15581	0.53	205	2791	0.35	7	0.17
ITEM 23	0.83	16.07	0.00	0.44 A	0.11	0.37	0.10	15593	0.51	208	2799	0.41	9	0.10
ITEM 24	0.84	14.39	0.00	0.42 A	0.11	0.34	0.10	15557	0.58	208	2790	0.46	9	0.11
ITEM 25	0.89	6.02	0.01	0.27 A	0.11	0.23	0.10	15376	0.52	208	2754	0.40	9	0.11
ITEM 26	1.44	63.22	0.00	-0.86 A	0.11	-0.74	0.10	15559	0.55	205	2769	0.33	7	0.22
ITEM 27	1.01	0.04	0.85	-0.02 A	0.11	0.00	0.10	15517	0.57	221	2750	0.40	15	0.17
ITEM 28	1.15	8.07	0.00	-0.32 A	0.11	-0.29	0.11	15496	0.48	208	2731	0.32	9	0.16
ITEM 29	0.77	30.66	0.00	0.62 A	0.11	0.55	0.11	15530	0.38	208	2751	0.33	9	0.05
ITEM 30	0.84	11.29	0.00	0.42 A	0.12	0.42	0.12	15472	0.26	208	2733	0.23	9	0.04

Source: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988: Base Year Survey .

Appendix B-4--(continued)

Differential Item Functioning (DIF), History/Citizenship/Geography

MANTEL-HAENSZEL ODDS-RATIO AND OTHER STATISTICS, NUMBER OF TABLES = 30

		NO. LEVELS	LEVEL 1	LEVEL 2
GROUP VARIABLE:	RACE	2	WHITE (REFERENCE)	AM IND (FOCAL)
RESPONSE VARIABLE:	ITEMSCOR	2	RIGHT	WRONG
STRATIFYING VARIABLE:	# RIGHT	31		

	MH ODDS RATIO	MH CHI-SQUARE	PROB > CHI-SQ	MH D-DIF		STD ERR MH D-DIF	STDZD D-DIF	STD ERR STD D-DIF	REFERENCE N	REFERENCE P+	REFERENCE NO*	FOCAL N	FOCAL P+	FOCAL NO*	IMPACT
ITEM 1	1.24	2.24	0.14	-0.51	A	0.33	-0.42	0.30	15457	0.85	208	299	0.69	0	0.16
ITEM 2	0.99	0.00	0.99	0.02	A	0.34	0.01	0.29	15668	0.82	208	306	0.66	0	0.16
ITEM 3	0.89	0.32	0.57	0.28	A	0.43	0.20	0.36	15677	0.93	2114	307	0.83	9	0.10
ITEM 4	1.21	1.97	0.16	-0.46	A	0.31	-0.35	0.28	15628	0.76	228	304	0.56	2	0.20
ITEM 5	1.09	0.22	0.64	-0.20	A	0.37	-0.17	0.34	15581	0.90	623	298	0.79	1	0.11
ITEM 6	0.88	0.48	0.49	0.29	A	0.38	0.25	0.35	15595	0.87	208	298	0.81	0	0.07
ITEM 7	1.50	4.29	0.04	-0.95	A	0.45	-0.63	0.36	15594	0.95	1955	299	0.82	7	0.13
ITEM 8	1.09	0.19	0.66	-0.21	A	0.40	-0.17	0.36	15583	0.92	837	299	0.82	2	0.10
ITEM 9	2.13	15.02	0.00	-1.78	C	0.46	-1.14	0.35	15596	0.95	2196	299	0.81	7	0.15
ITEM 10	0.86	1.28	0.26	0.35	A	0.30	0.32	0.28	15638	0.73	208	303	0.62	0	0.10
ITEM 11	1.05	0.10	0.75	-0.12	A	0.31	-0.09	0.28	15637	0.65	208	303	0.45	0	0.20
ITEM 12	0.93	0.28	0.60	0.18	A	0.30	0.16	0.28	15623	0.59	208	303	0.44	0	0.15
ITEM 13	0.77	3.77	0.05	0.61	A	0.30	0.50	0.27	15560	0.63	208	302	0.50	0	0.13
ITEM 14	0.86	1.24	0.27	0.35	A	0.30	0.33	0.29	15541	0.44	240	304	0.34	4	0.10
ITEM 15	0.91	0.48	0.49	0.23	A	0.31	0.21	0.28	15654	0.52	208	308	0.36	0	0.17
ITEM 16	0.84	1.76	0.19	0.40	A	0.29	0.38	0.28	15643	0.48	208	307	0.38	0	0.10
ITEM 17	1.30	3.00	0.08	-0.61	A	0.34	-0.47	0.30	15634	0.87	208	305	0.70	0	0.17
ITEM 18	1.33	3.92	0.05	-0.67	A	0.32	-0.53	0.28	15653	0.82	208	306	0.63	0	0.19
ITEM 19	0.92	0.23	0.63	0.19	A	0.34	0.14	0.28	15630	0.81	623	306	0.63	1	0.18
ITEM 20	1.27	2.99	0.08	-0.55	A	0.31	-0.44	0.27	15609	0.72	208	306	0.50	0	0.22
ITEM 21	0.99	0.00	0.99	0.02	A	0.30	0.02	0.28	15590	0.72	208	304	0.55	0	0.17
ITEM 22	0.80	2.70	0.10	0.52	A	0.30	0.44	0.28	15581	0.53	208	303	0.40	0	0.13
ITEM 23	0.90	0.65	0.42	0.26	A	0.30	0.23	0.28	15593	0.51	208	306	0.39	0	0.13
ITEM 24	0.91	0.50	0.48	0.23	A	0.30	0.20	0.28	15557	0.58	208	303	0.43	0	0.15
ITEM 25	0.92	0.32	0.57	0.19	A	0.30	0.18	0.28	15376	0.52	220	300	0.39	1	0.13
ITEM 26	1.15	1.10	0.29	-0.34	A	0.30	-0.30	0.29	15559	0.55	221	303	0.36	1	0.19
ITEM 27	1.11	0.56	0.46	-0.25	A	0.31	-0.20	0.29	15517	0.57	221	301	0.36	1	0.21
ITEM 28	0.98	0.00	0.96	0.04	A	0.30	0.03	0.29	15496	0.48	208	303	0.34	0	0.14
ITEM 29	0.73	5.89	0.02	0.75	A	0.30	0.70	0.29	15530	0.38	208	301	0.34	0	0.04
ITEM 30	0.79	2.65	0.10	0.57	A	0.33	0.54	0.33	15472	0.26	208	301	0.23	0	0.03

Source: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988: Base Year Survey .

Appendix B-4--(continued)

Differential Item Functioning (DIF), History/Citizenship/Geography

MANTEL-HAENSZEL ODDS-RATIO AND OTHER STATISTICS, NUMBER OF TABLES = 30

		NO. LEVELS	LEVEL 1		LEVEL 2	
GROUP VARIABLE:	SEX	2	MALE	(REFERENCE)	FEMALE	(FOCAL)
RESPONSE VARIABLE:	ITEMSCOR	2	RIGHT		WRONG	
STRATIFYING VARIABLE:	# RIGHT	31				

	MH ODDS RATIO	MH CHI-SQUARE	PROB > CHI-SQ	MH D-DIF		STD ERR MH D-DIF	STDZD D-DIF	STD ERR STD D-DIF	REFERENCE N	REFERENCE P+	REFERENCE NO*	FOCAL N	FOCAL P+	FOCAL NO*	IMPACT
ITEM 1	0.97	0.63	0.43	0.07	A	0.09	0.06	0.08	11363	0.82	159	11584	0.82	95	0.01
ITEM 2	0.88	11.38	0.00	0.29	A	0.09	0.22	0.07	11586	0.78	160	11724	0.79	98	-0.01
ITEM 3	0.71	43.32	0.00	0.82	A	0.12	0.62	0.10	11585	0.89	1516	11735	0.92	1077	-0.02
ITEM 4	1.18	23.98	0.00	-0.38	A	0.08	-0.30	0.06	11563	0.72	159	11659	0.67	95	0.04
ITEM 5	1.15	10.49	0.00	-0.33	A	0.10	-0.27	0.09	11458	0.89	475	11663	0.87	275	0.02
ITEM 6	0.87	12.35	0.00	0.33	A	0.09	0.30	0.09	11459	0.84	159	11679	0.86	96	-0.01
ITEM 7	0.99	0.03	0.86	0.03	A	0.14	0.02	0.11	11469	0.92	1402	11662	0.92	955	0.00
ITEM 8	1.12	5.71	0.02	-0.27	A	0.11	-0.22	0.10	11458	0.90	604	11671	0.89	411	0.01
ITEM 9	1.09	1.95	0.16	-0.20	A	0.14	-0.14	0.11	11455	0.92	920	11683	0.92	591	0.00
ITEM 10	0.93	5.06	0.02	0.17	A	0.07	0.15	0.07	11548	0.71	159	11703	0.70	95	0.01
ITEM 11	1.29	65.62	0.00	-0.61	A	0.07	-0.45	0.06	11546	0.65	164	11687	0.57	100	0.08
ITEM 12	0.66	200.50	0.00	0.98	A	0.07	0.83	0.06	11527	0.54	160	11676	0.60	98	-0.06
ITEM 13	1.23	45.23	0.00	-0.49	A	0.07	-0.38	0.06	11497	0.63	165	11598	0.57	101	0.07
ITEM 14	1.08	6.93	0.01	-0.18	A	0.07	-0.16	0.06	11473	0.47	160	11620	0.43	98	0.04
ITEM 15	0.93	5.04	0.02	0.16	A	0.07	0.13	0.06	11562	0.49	160	11705	0.47	97	0.02
ITEM 16	1.00	0.02	0.90	0.01	A	0.07	0.02	0.06	11554	0.48	160	11688	0.45	97	0.03
ITEM 17	1.05	1.28	0.26	-0.11	A	0.10	-0.10	0.08	11553	0.85	159	11678	0.84	95	0.01
ITEM 18	0.98	0.26	0.61	0.04	A	0.08	0.03	0.07	11550	0.79	158	11708	0.78	95	0.01
ITEM 19	0.64	127.95	0.00	1.03	B	0.09	0.72	0.07	11540	0.76	479	11681	0.80	280	-0.04
ITEM 20	1.02	0.42	0.52	-0.05	A	0.07	-0.05	0.06	11526	0.68	163	11658	0.66	100	0.02
ITEM 21	2.19	580.92	0.00	-1.85	C	0.08	-1.48	0.06	11513	0.76	159	11639	0.60	97	0.15
ITEM 22	0.86	23.86	0.00	0.35	A	0.07	0.28	0.06	11494	0.50	159	11628	0.49	97	0.01
ITEM 23	0.94	4.26	0.04	0.14	A	0.07	0.14	0.06	11499	0.50	158	11651	0.48	95	0.02
ITEM 24	0.90	12.59	0.00	0.25	A	0.07	0.22	0.06	11475	0.55	163	11625	0.55	100	0.00
ITEM 25	0.79	62.11	0.00	0.54	A	0.07	0.47	0.06	11371	0.48	164	11441	0.50	100	-0.02
ITEM 26	1.13	18.18	0.00	-0.30	A	0.07	-0.25	0.06	11457	0.53	159	11607	0.47	97	0.06
ITEM 27	0.94	3.85	0.05	0.14	A	0.07	0.12	0.06	11439	0.54	163	11541	0.52	98	0.02
ITEM 28	1.31	83.55	0.00	-0.64	A	0.07	-0.52	0.06	11411	0.48	163	11507	0.40	100	0.09
ITEM 29	0.88	17.86	0.00	0.29	A	0.07	0.24	0.06	11419	0.37	162	11565	0.36	100	0.00
ITEM 30	1.07	4.00	0.05	-0.15	A	0.08	-0.17	0.07	11370	0.27	162	11520	0.24	100	0.04

Source: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988: Base Year Survey .

APPENDIX C

ITEM PARAMETERS

ITEM PARAMETERS FOR READING TEST

ITEM NUMBER	A	S.E.	B	S.E	C	S.E
ITEM 1	0.5250	(0.018)	-4.8212	(0.162)	0.1443	(0.031)
ITEM 2	0.7529	(0.016)	-1.9058	(0.039)	0.1443	(0.011)
ITEM 3	0.8132	(0.017)	-1.5510	(0.032)	0.1443	(0.010)
ITEM 4	0.8621	(0.017)	-0.2266	(0.018)	0.0992	(0.007)
ITEM 5	1.3226	(0.029)	0.1287	(0.014)	0.2013	(0.006)
ITEM 6	0.9888	(0.021)	-0.1285	(0.019)	0.1954	(0.008)
ITEM 7	1.0526	(0.024)	0.5996	(0.014)	0.1267	(0.005)
ITEM 8	0.9751	(0.019)	0.1704	(0.015)	0.1026	(0.006)
ITEM 9	0.7863	(0.022)	0.0476	(0.029)	0.2993	(0.009)
ITEM 10	0.3534	(0.013)	1.7075	(0.063)	0.1834	(0.010)
ITEM 11	0.9849	(0.022)	-0.0339	(0.019)	0.2075	(0.008)
ITEM 12	1.3770	(0.026)	-0.6228	(0.015)	0.1700	(0.007)
ITEM 13	1.5527	(0.045)	0.6267	(0.014)	0.3172	(0.005)
ITEM 14	1.5068	(0.035)	0.4419	(0.012)	0.2078	(0.005)
ITEM 15	1.1584	(0.023)	0.2694	(0.013)	0.1083	(0.005)
ITEM 16	1.3549	(0.028)	-0.7676	(0.018)	0.2425	(0.009)
ITEM 17	1.8182	(0.043)	0.3088	(0.011)	0.2589	(0.005)
ITEM 18	0.7303	(0.021)	0.4045	(0.027)	0.2391	(0.009)
ITEM 19	1.1892	(0.026)	-0.1504	(0.017)	0.2270	(0.008)
ITEM 20	1.1135	(0.027)	-0.3595	(0.022)	0.3091	(0.009)
ITEM 21	1.2877	(0.033)	0.1028	(0.018)	0.3176	(0.007)
MEAN	1.0717		-0.2743		0.2022	
S.D	0.3473		1.2565		0.0693	

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988: Base Year Survey.

ITEM PARAMETERS FOR MATHEMATICS TEST

ITEM NUMBER	A	S.E.	B	S.E	C	S.E
ITEM 1	1.2329	(0.024)	-0.6117	(0.018)	0.1866	(0.009)
ITEM 2	0.9232	(0.021)	0.2578	(0.019)	0.1534	(0.007)
ITEM 3	1.0972	(0.055)	1.4866	(0.028)	0.4083	(0.005)
ITEM 4	1.3225	(0.029)	0.3042	(0.013)	0.1890	(0.006)
ITEM 5	1.3625	(0.030)	0.2080	(0.014)	0.2041	(0.006)
ITEM 6	1.2673	(0.041)	0.9306	(0.017)	0.3048	(0.005)
ITEM 7	1.4483	(0.030)	0.4492	(0.011)	0.1320	(0.005)
ITEM 8	1.2523	(0.031)	0.7607	(0.013)	0.1560	(0.005)
ITEM 9	1.6205	(0.045)	0.7538	(0.012)	0.2732	(0.005)
ITEM 10	1.2382	(0.030)	0.6206	(0.013)	0.1696	(0.005)
ITEM 11	1.1173	(0.030)	0.8894	(0.015)	0.1651	(0.005)
ITEM 12	1.0766	(0.022)	0.3406	(0.014)	0.1118	(0.006)
ITEM 13	1.3096	(0.026)	0.0876	(0.013)	0.1555	(0.006)
ITEM 14	1.3019	(0.027)	0.1736	(0.013)	0.1539	(0.006)
ITEM 15	0.7174	(0.019)	-0.6095	(0.041)	0.2684	(0.014)
ITEM 16	0.5423	(0.012)	-1.6847	(0.051)	0.1049	(0.015)
ITEM 17	0.4751	(0.012)	-1.1686	(0.054)	0.1049	(0.015)
ITEM 18	1.5441	(0.035)	0.3016	(0.012)	0.2372	(0.006)
ITEM 19	0.7709	(0.015)	-1.4074	(0.032)	0.1049	(0.012)
ITEM 20	0.6127	(0.013)	-1.7501	(0.045)	0.1049	(0.014)
ITEM 21	0.6777	(0.013)	-0.8586	(0.029)	0.0761	(0.010)
ITEM 22	1.1909	(0.020)	-0.6475	(0.015)	0.0826	(0.007)
ITEM 23	0.4309	(0.012)	0.8505	(0.058)	0.1049	(0.015)
ITEM 24	0.7683	(0.018)	-0.1930	(0.027)	0.1552	(0.010)
ITEM 25	1.0249	(0.020)	-0.4229	(0.020)	0.1484	(0.009)
ITEM 26	1.3040	(0.033)	0.0725	(0.018)	0.3265	(0.008)
ITEM 27	1.7307	(0.032)	-0.2009	(0.011)	0.1534	(0.006)
ITEM 28	0.8015	(0.017)	-0.1632	(0.022)	0.1053	(0.009)
ITEM 29	1.0219	(0.021)	0.0455	(0.016)	0.1194	(0.007)
ITEM 30	0.7250	(0.019)	0.2235	(0.027)	0.1680	(0.010)
ITEM 31	1.2122	(0.024)	-0.1408	(0.016)	0.1699	(0.007)
ITEM 32	0.9630	(0.026)	-0.1005	(0.028)	0.3407	(0.010)
ITEM 33	0.4860	(0.025)	1.3687	(0.051)	0.2753	(0.012)
ITEM 34	1.5186	(0.037)	0.3902	(0.013)	0.2741	(0.006)
ITEM 35	0.7955	(0.024)	0.2805	(0.029)	0.2753	(0.010)
ITEM 36	1.3104	(0.030)	0.5704	(0.012)	0.1555	(0.005)
ITEM 37	1.0067	(0.018)	0.1768	(0.012)	0.0369	(0.005)
ITEM 38	0.8602	(0.042)	1.5293	(0.031)	0.3254	(0.006)
ITEM 39	2.1037	(0.045)	0.5591	(0.008)	0.1487	(0.004)
ITEM 40	1.7370	(0.042)	0.9381	(0.010)	0.1233	(0.003)
MEAN	1.0976		0.0727		0.1813	
S.D	0.3785		0.7758		0.0835	

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988: Base Year Survey.

ITEM PARAMETERS FOR SCIENCE TEST

ITEM NUMBER	A	S.E.	B	S.E	C	S.E
ITEM 1	1.2929	(0.034)	-0.0888	(0.021)	0.3800	(0.008)
ITEM 2	0.5494	(0.012)	-1.6620	(0.045)	0.0931	(0.013)
ITEM 3	0.6050	(0.016)	-0.3815	(0.043)	0.2053	(0.013)
ITEM 4	0.6218	(0.020)	-0.1582	(0.049)	0.3188	(0.014)
ITEM 5	1.2829	(0.018)	-0.9936	(0.011)	0.0046	(0.003)
ITEM 6	1.0064	(0.015)	-1.1211	(0.014)	0.0069	(0.003)
ITEM 7	0.5666	(0.014)	-0.5728	(0.042)	0.1519	(0.013)
ITEM 8	0.7106	(0.023)	0.2856	(0.033)	0.2672	(0.010)
ITEM 9	0.5484	(0.012)	0.6843	(0.037)	0.0931	(0.011)
ITEM 10	1.2138	(0.032)	0.3911	(0.017)	0.2802	(0.007)
ITEM 11	0.6029	(0.025)	0.9040	(0.037)	0.2653	(0.010)
ITEM 12	0.8157	(0.018)	-0.5085	(0.028)	0.1704	(0.011)
ITEM 13	0.6516	(0.014)	-1.0218	(0.039)	0.1519	(0.013)
ITEM 14	1.7614	(0.036)	0.1574	(0.010)	0.1937	(0.005)
ITEM 15	0.5516	(0.018)	0.8469	(0.030)	0.1135	(0.009)
ITEM 16	1.1648	(0.041)	0.9907	(0.019)	0.3255	(0.006)
ITEM 17	1.5097	(0.042)	0.8177	(0.013)	0.2475	(0.005)
ITEM 18	1.2889	(0.034)	0.6395	(0.014)	0.2323	(0.006)
ITEM 19	1.3258	(0.037)	0.7987	(0.014)	0.2417	(0.005)
ITEM 20	1.6855	(0.066)	1.2473	(0.016)	0.3351	(0.004)
ITEM 21	1.3803	(0.050)	1.1371	(0.017)	0.3160	(0.005)
ITEM 22	0.8041	(0.035)	1.4299	(0.028)	0.2441	(0.007)
ITEM 23	1.0786	(0.061)	1.7891	(0.035)	0.3458	(0.005)
ITEM 24	0.8942	(0.022)	0.8113	(0.015	0.0765	(0.005)
ITEM 25	0.6996	(0.032)	2.0071	(0.042)	0.1121	(0.005)
MEAN	0.9845		0.2824		0.2069	
S.D	0.3749		0.9500		0.1040	

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988: Base Year Survey.

ITEM PARAMETERS FOR HISTORY/CITIZENSHIP/GEOGRAPHY TEST

ITEM NUMBER	A	S.E.	B	S.E	C	S.E
ITEM 1	1.0496	(0.030)	-0.5444	(0.035)	0.4565	(0.012)
ITEM 2	0.9833	(0.021)	-0.8964	(0.029)	0.2195	(0.012)
ITEM 3	1.6649	(0.044)	-1.3435	(0.025)	0.3644	(0.013)
ITEM 4	1.0102	(0.023)	-0.3776	(0.024)	0.2367	(0.010)
ITEM 5	1.1296	(0.031)	-1.0224	(0.038)	0.4635	(0.013)
ITEM 6	0.5205	(0.017)	-1.6335	(0.094)	0.3680	(0.023)
ITEM 7	1.5133	(0.033)	-1.8517	(0.021)	0.0826	(0.011)
ITEM 8	0.9790	(0.022)	-1.7132	(0.036)	0.2097	(0.016)
ITEM 9	1.5849	(0.035)	-1.8688	(0.020)	0.0762	(0.010)
ITEM 10	1.1069	(0.036)	0.2149	(0.027)	0.4689	(0.008)
ITEM 11	2.0744	(0.049)	0.1959	(0.011)	0.2964	(0.006)
ITEM 12	0.7068	(0.020)	0.1729	(0.030)	0.1911	(0.010)
ITEM 13	1.4423	(0.036)	0.2593	(0.015)	0.3025	(0.006)
ITEM 14	0.9478	(0.034)	1.0496	(0.021)	0.2660	(0.006)
ITEM 15	1.3145	(0.031)	0.4760	(0.013)	0.2020	(0.006)
ITEM 16	1.5454	(0.047)	0.8897	(0.014)	0.3017	(0.005)
ITEM 17	0.8238	(0.018)	-1.4562	(0.039)	0.1947	(0.016)
ITEM 18	0.9370	(0.025)	-0.6494	(0.036)	0.3659	(0.013)
ITEM 19	1.6059	(0.034)	-0.6313	(0.017)	0.2572	(0.009)
ITEM 20	0.8968	(0.021)	-0.2790	(0.027)	0.2226	(0.010)
ITEM 21	1.1929	(0.030)	-0.0569	(0.021)	0.3294	(0.008)
ITEM 22	1.4767	(0.037)	0.5534	(0.013)	0.2538	(0.005)
ITEM 23	1.2290	(0.037)	0.7582	(0.016)	0.2912	(0.006)
ITEM 24	0.7872	(0.021)	0.2554	(0.025)	0.1891	(0.009)
ITEM 25	0.8587	(0.028)	0.7691	(0.023)	0.2539	(0.008)
ITEM 26	1.2166	(0.033)	0.6286	(0.016)	0.2620	(0.006)
ITEM 27	1.1746	(0.027)	0.2807	(0.015)	0.1878	(0.007)
ITEM 28	1.8998	(0.055)	0.8826	(0.011)	0.2814	(0.004)
ITEM 29	1.4052	(0.053)	1.3309	(0.017)	0.2611	(0.004)
ITEM 30	2.2371	(0.089)	1.5372	(0.013)	0.1902	(0.003)
MEAN	1.2438		-0.1357		0.2682	
S.D	0.3974		0.9715		0.0941	

SOURCE: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988: Base Year Survey.

APPENDIX D

TEST INFORMATION FUNCTIONS

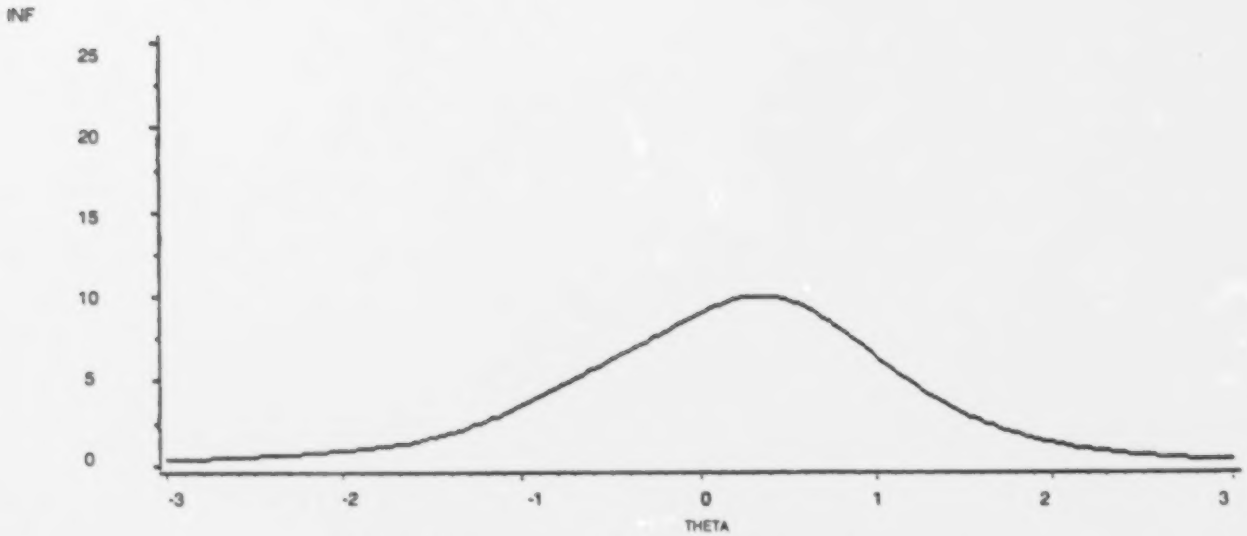
APPENDIX D

Test Information Functions

Appendix D presents the test information functions for the 8th Grade test forms. The test information functions can be interpreted as a plot of the reciprocal of the square of the standard error of measurement for all values of theta. In general, information functions of 1.0 and higher are considered quite acceptable. Over 90% of the students' scores are in the theta range that meets this criterion on all four tests. The information functions for Science and History/Citizenship/Geography are less peaked and have broad band measurement properties. Reading and Mathematics are slightly more peaked, with the best measurement slightly above the mean.

APPENDIX D-1

NELS:88 Grade 8 Reading Test
21 Items
Test Information Function

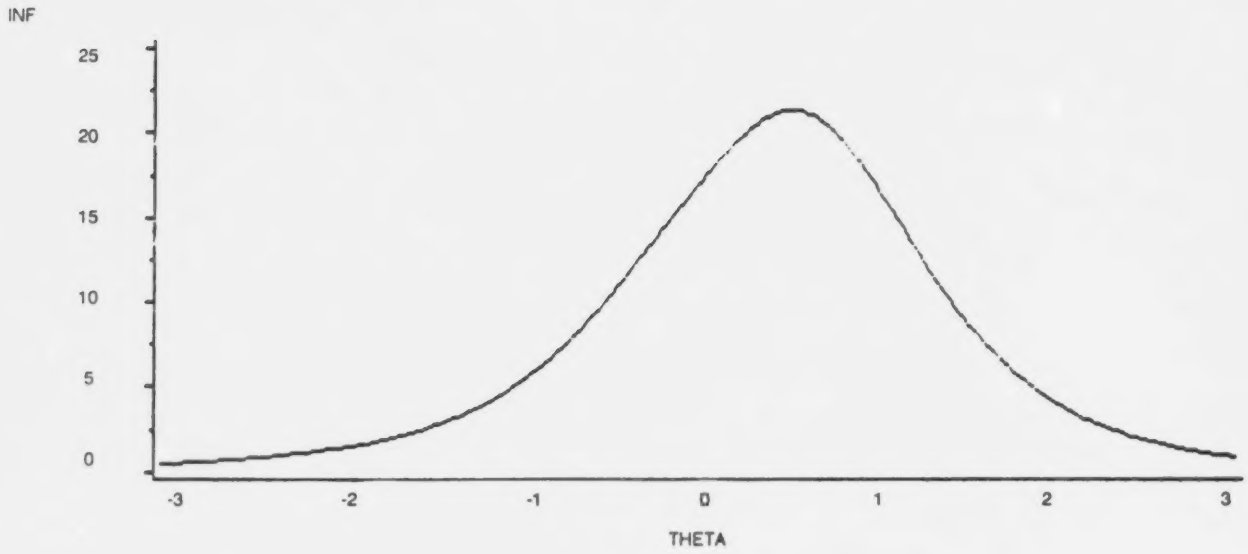


Information function - reciprocal of square of standard error of measurement.

Source: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988: Base Year Survey.

APPENDIX D-2

NELS:88 Grade 8 Mathematics Test
40 Items
Test Information Function

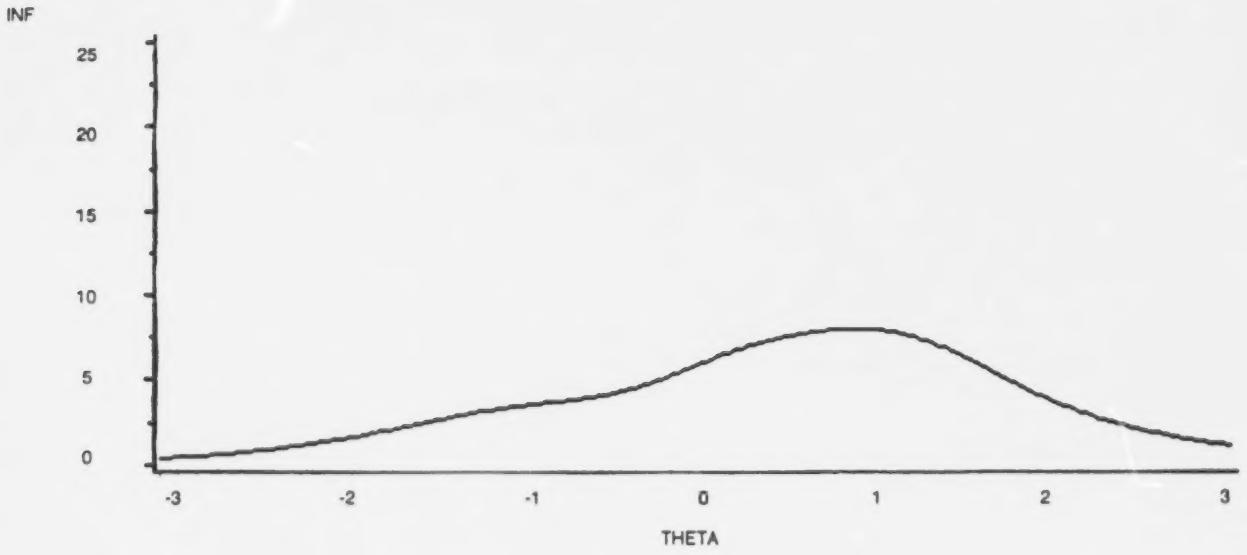


Information function - reciprocal of square of standard error of measurement.

Source: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988: Base Year Survey.

APPENDIX D-3

NELS:88 Grade 8 Science Test
25 Items
Test Information Function

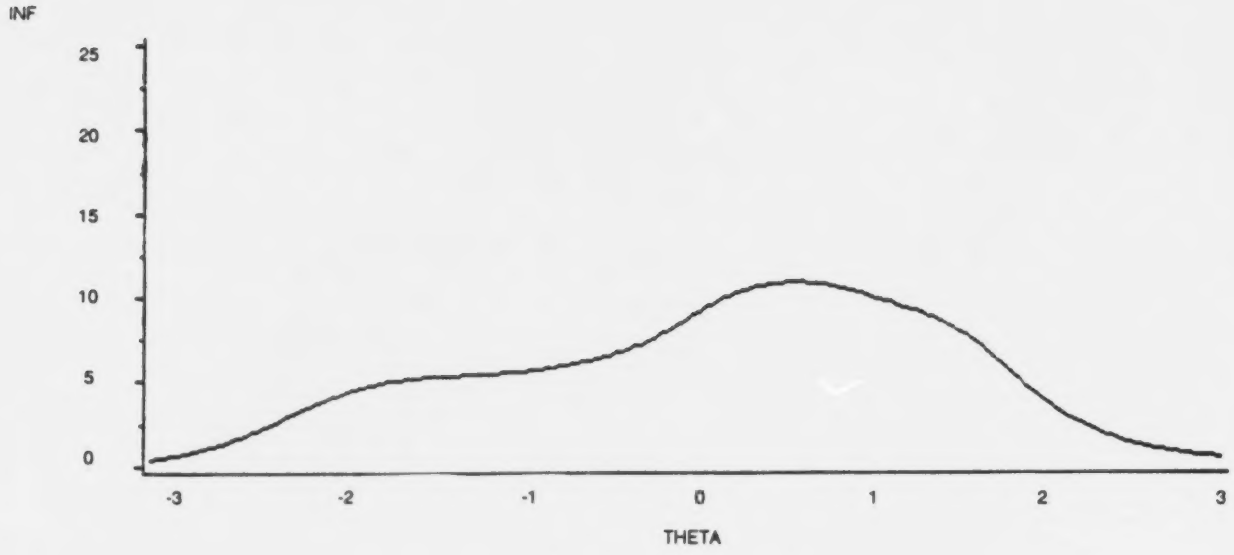


Information function - reciprocal of square of standard error of measurement.

Source: U.S. Department of Education, National Center for Education Statistics, National Education Longitudinal Study of 1988: Base Year Survey.

APPENDIX D-4

NELS:88 Grade 8 History Test
30 Items
Test Information Function



Information function - reciprocal of square of standard error of measurement.

Source: U.S. Department of Education, National Center for Education Statistics, "National Education Longitudinal Study of 1988: Base Year Survey.

APPENDIX E

DESCRIPTION OF INDIVIDUAL ITEMS

APPENDIX E-1

Description of Reading Comprehension Items

Item	Content	Process	# Options	Source	Description of Reading Passages and Items
				Reading	Passage 1: A fable containing dialogue between two characters.
1	Literary	Repro-Detail	5	NAEP-R	Identify the objective of a character's course of action
2	Literary	Repro-Detail	5	NELS	Identify a character's assumption in planning his actions
3	Literary	Repro-Detail	5	NAEP-R	Identify the reason the character's plan didn't work
4	Literary	Inference/Eval	5	NELS	Choose which personality trait is suggested by the story
5	Literary	Inference/Eval	5	NELS	Choose the adage that best fits the lesson to be learned
				Reading	Passage 2: A paragraph relating events in geologic time and evolution to the span of a year.
6	Science	Repro-Detail	5	NELS	Demonstrate understanding of the time-line metaphor
7	Science	Inference/Eval	5	HSB	Choose the event the author seems least certain about
8	Science	Comprehension	5	NELS	Relate two events using the time-line
				Reading	Passage 3: A metaphorical poem consisting of parallels between the author's emotional crisis and a writing assignment
9	Poetry	Comprehension	4	3IBR-R	Identify the tension or conflict implied in the poem
10	Poetry	Inference/Eval	4	3IBR-R	Infer the meaning of a metaphor from the context of the line
11	Poetry	Inference/Eval	4	3IBR-R	Evaluate personality traits suggested by the poem
12	Poetry	Inference/Eval	4	3IBR-R	Choose the mood suggested by the tone of a phrase
13	Poetry	Inference/Eval	4	3IBR-R	Identify the author's state of mind
14	Poetry	Inference/Eval	4	NELS	Identify an example of personification

APPENDIX E-1 (Continued)

Description of Reading Comprehension Items

Item	Content	Process	# Options	Source	Description of Reading Passages and Items
				Reading	Passage 4: A short biography of a Black musician.
15	Biography	Comprehension	4	3IBR	Evaluate the main purpose of the passage
16	Biography	Inference/Eval	4	3IBR	Define the meaning of a phrase
17	Biography	Inference/Eval	4	3IBR	Evaluate the tone of a character's remark in context
18	Biography	Inference/Eval	4	3IBR	Choose a statement supported by evidence in passage
				Reading	Passage 5: A short essay on the experiences of pioneer women in the United States.
19	Literary	Inference/Eval	4		NELS Identify author's reason for a quote from a diary
20	Literary	Inference/Eval	4	3IBR	Identify author's attitude toward pioneer women
21	Literary	Inference/Eval	4	NELS	Explain reason for a specified assumption

Notes: The designation "-R" indicates that the item has been revised from the original. 3IBR is the form code designation for a test previously used in an ETS testing program.

APPENDIX E-2

Description of Mathematics Items

Item	Content	Process	# Options	Source	Item Description
1	Algebra	Skill/Knowledge	4	HSB	Compare 2 algebraic expressions, given values of variables
2	Data/Prob	Und/Comp	4	HSB	Compare two numbers read from a graph
3	Data/Prob	Skill/Knowledge	4	HSB	Read two numbers from a graph and perform an operation with them
4	Algebra	Und/Comp	4	HSB	Compare two algebraic expressions, given a relationship
5	Arithmetic	Skill/Knowledge	4	HSB	Perform an arithmetic operation and compare result with a number
6	Adv. Topics	Skill/Knowledge	4	HSB	Determine coordinates of points on a graph, perform an operation
7	Algebra	Und/Comp	4	HSB	Compare two algebraic expressions
8	Arithmetic	Skill/Knowledge	4	HSB	Perform an arithmetic operation, compare result with a number
9	Arithmetic	Skill/Knowledge	4	HSB	Perform an arithmetic operation, compare result with a number
10	Arithmetic	Und/Comp	4	HSB	Compare statements about locations on two number lines
11	Geometry	Und/Comp	4	HSB	Compare length of line segments illustrated in a diagram
12	Arithmetic	Skill/Knowledge	4	HSB	Compare expressions involving mult. and division of integers
13	Arithmetic	Skill/Knowledge	4	HSB	Compare an integer with an expression using division of decimals
14	Algebra	Und/Comp	4	HSB	Compare expressions, given information containing exponents
15	Algebra	Skill/Knowledge	4	HSB	Compare expressions, requiring solution of simple equations
16	Arithmetic	Skill/Knowledge	4	HSB	Compare two quantities of money expressed differently
17	Arithmetic	Skill/Knowledge	4	HSB	Compare two simple arithmetic expressions involving division
18	Arithmetic	Skill/Knowledge	4	NELS	Compare two simple arithmetic expressions involving division
19	Arithmetic	Skill/Knowledge	4	NELS	Compare two simple arithmetic expressions involving multiplic.
20	Arithmetic	Und/Comp	4	NAEP	Set up a simple equation that is the solution of a word problem
21	Data/Prob	Und/Comp	5	NAEP	Estimate a probability that is the solution of a word problem
22	Arithmetic	Skill/Knowledge	4	NAEP	Determine the greatest of 4 decimal numbers
23	Arithmetic	Problem Solving	4	NAEP	Determine the smallest of 4 fractions in a word problem
24	Data/Prob	Und/Comp	4	NAEP	Choose verbal description of a prob. that doesn't match diagram
25	Geometry	Skill/Knowledge	5	NAEP	Determine the length of a line segment in a diagram
26	Algebra	Und/Comp	4	NAEP	Evaluate a relationship given statements about the variables
27	Algebra	Und/Comp	4	NAEP	Find an algebraic expression odd or even given fact about var.
28	Arithmetic	Problem Solving	4	NAEP	Solve a word problem requiring logical inference

APPENDIX E-2 (Continued)

Description of Mathematics Items

Item	Content	Process	# Options	Source	Item Description
29	Algebra	Und/Comp	5	NAEP	Solve a word problem whose answer is an algebraic expression
30	Arithmetic	Problem Solving	4	NAEP	Solve a word problem using multiplication or factoring
31	Arithmetic	Und/Comp	4	NAEP	Choose which decimal number is between two other numbers
32	Arithmetic	Und/Comp	4	NAEP	Choose points on a number line that include a specified decimal
33	Arithmetic	Und/Comp	5	NAEP	Estimate a number using a percentage indicated in a diagram
34	Algebra	Skill/Knowledge	4	NAEP	Solve a simple algebraic equation
35	Adv. Topics	Problem Solving	4	NAEP	Evaluate statements inferred from a word problem with a fraction
36	Arithmetic	Und/Comp	4	NAEP	Choose which expression is different from a specified percentage
37	Geometry	Und/Comp	4	NAEP	Solve a word problem requiring logical inference
38	Geometry	Und/Comp	4	NAEP	Evaluate statements referring to area and diagonal of a diagram
39	Algebra	Und/Comp	4	NAEP	Supply number that completes an algebraic equation correctly
40	Algebra	Skill/Knowledge	5	NAEP	Simplify an algebraic expression

APPENDIX E-3

Description of Science Items

Item	Content	Process	# Options	Source	Item Description
1	Earth Sci	Problem Solving	4	NAEP	Infer geologic history from facts about limestone deposits
2	Earth Sci	Decl Knowledge	5	NAEP	Identify components of solar system
3	Chemistry	Und/Comp	4	NAEP	Read a graph depicting solubility of chemicals
4	Sci Method	Problem Solving	4	NAEP	Choose an improvement for an experiment on mice
5	Earth Sci	Decl Knowledge	5	HSB	Choose a statement about source of moon's light
6	Life Sci	Decl Knowledge	5	HSB	Identify the example of a simple reflex
7	Earth Sci	Und/Comp	4	NAEP	Choose viable way of communicating on the moon
8	Earth Sci	Decl Knowledge	4	NAEP	Select statement about position of sun, moon, earth in diagram
9	Life Sci	Decl Knowledge	5	NELS	Identify source of oxygen in ocean water
10	Chemistry	Decl Knowledge	4	NAEP	Choose the property used to classify a list of substances
11	Chemistry	Comprehension	4	NAEP	Explain lower freezing temperature of ocean water
12	Earth Sci	Decl Knowledge	5	HSB	Answer question about the earth's orbit
13	Life Sci	Problem Solving	4	NAEP	Infer use of oxygen from description of condition of aquarium
14	Chemistry	Problem Solving	5	HSB	Estimate temperature of a mixture
15	Life Sci	Decl Knowledge	4	NAEP	Select a statement about the process of respiration
16	Life Sci	Und/Comp	4	NAEP	Read a graph depicting digestion of a protein by an enzyme
17	Life Sci	Und/Comp	4	NAEP	Explain location of marine algae
18	Earth Sci	Decl Knowledge	4	NAEP	Choose best indication of an approaching storm
19	Chemistry	Decl Knowledge	4	NAEP	Choose the alternative that is NOT a chemical change
20	Chemistry	Problem Solving	4	NAEP	Infer statement from results of an experiment using a filter
21	Earth Sci	Und/Comp	4	NAEP	Explain reason for late afternoon breeze from the ocean
22	Life Sci	Problem Solving	4	NAEP	Select basis for a statement about a food chain
23	Chemistry	Problem Solving	4	NAEP	Interpret symbols describing a chemical reaction
24	Sci Method	Und/Comp	5	HSB	Differentiate statements based on a model or an observation
25	Life Sci	Problem Solving	5	HSB	Describe color of offspring from a guinea pig cross

APPENDIX E-4

Description of History/Citizenship/Geography Items

Item	Content	# Options	Source	Item Description
1	Geography	4	NAEP	Historical time line indicating how people have obtained food
2	History	4	NAEP	Definition of a Civil War era institution
3	Citizenship	4	NAEP	Identify a phrase that is NOT a constitutional right
4	History	4	NAEP	Identify a historically important manufacturing technique
5	Citizenship	2	NAEP	Indicate whether an action is legal or not legal
6	Citizenship	2	NAEP	Indicate whether an action is legal or not legal
7	Citizenship	2	NAEP	Indicate whether an action is legal or not legal
8	Citizenship	2	NAEP	Indicate whether an action is legal or not legal
9	Citizenship	2	NAEP	Indicate whether an action is legal or not legal
10	History	4	NAEP	Identify source of guarantees of specific freedoms
11	History	4	NAEP	Identify an important historical document
12	Geography	4	NAEP	Choose best explanation for facts about diet of most people in the world
13	History	4	NELS	Identify the president affected by an important historical event
14	History	4	NAEP	Complete a statement about immigration patterns
15	Citizenship	5	NAEP	Choose the correct option concerning the U.S. Congress
16	Citizenship	5	NAEP	Choose the correct option concerning the U.S. Congress
17	History	4	NAEP	Identify the organization described
18	History	4	NELS	Identify the author of an important historical document
19	Citizenship	5	NAEP	Identify one of the purposes of an important historical document
20	History	4	NAEP	Identify a new feature of U.S. homes at a specified time period
21	History	4	NAEP	Identify the location and time of an important historical event
22	Citizenship	4	NAEP	Identify an underlying concept in the organization of the government
23	Citizenship	4	NAEP	Identify the branch of government that has a specified authority
24	Citizenship	4	HSB	Identify the principle exemplified by a specified right
25	History	4	NAEP	Identify the meaning of a specified Supreme Court decision
26	Geography	4	NAEP	Choose the option that identifies patterns of settlement
27	History	4	NAEP	Identify the purpose of a specified law
28	History	4	NAEP	Identify a factor that influenced population movement at a given time
29	History	4	NAEP	Identify the principal effect of specified legal requirements
30	Citizenship	4	HSB	Identify the principle exemplified by a specified legal requirement

APPENDIX F

INTERCORRELATIONS OF TESTLETS

APPENDIX F
Intercorrelations of Testlets

	READ-LIT	READ-SCI	READ-POE	READ-BIO	READ-HST	ARITH	ALGEBRA	GEOMETRY
READ-LIT	1.00	0.46	0.48	0.46	0.41	0.47	0.46	0.17
READ-SCI	0.46	1.00	0.48	0.46	0.40	0.54	0.51	0.20
READ-POE	0.48	0.48	1.00	0.53	0.47	0.54	0.53	0.21
READ-BIO	0.46	0.46	0.53	1.00	0.52	0.51	0.51	0.21
READ-HST	0.41	0.40	0.47	0.52	1.00	0.48	0.46	0.20
ARITH	0.47	0.54	0.54	0.51	0.48	1.00	0.80	0.32
ALGEBRA	0.46	0.51	0.53	0.51	0.46	0.80	1.00	0.32
GEOMETRY	0.17	0.20	0.21	0.21	0.20	0.32	0.32	1.00
PROBILTY	0.31	0.34	0.32	0.31	0.29	0.49	0.46	0.19
EARTHSCI	0.42	0.44	0.45	0.43	0.40	0.55	0.51	0.22
LIFE SCI	0.42	0.43	0.47	0.45	0.40	0.54	0.52	0.20
CHEMISTR	0.35	0.40	0.40	0.38	0.36	0.54	0.52	0.23
SCI METH	0.29	0.30	0.33	0.31	0.29	0.36	0.34	0.14
HISTORY	0.47	0.48	0.50	0.49	0.44	0.56	0.54	0.23
CIT/GOVT	0.47	0.47	0.50	0.50	0.45	0.58	0.56	0.23
GEOG/EC	0.42	0.43	0.45	0.45	0.42	0.53	0.51	0.22

	PROBILTY	EARTHSCI	LIFE SCI	CHEMISTR	SCI METH	HISTORY	CIT/GOVT	GEOG/EC
READ-LIT	0.31	0.42	0.42	0.35	0.29	0.47	0.47	0.42
READ-SCI	0.34	0.44	0.43	0.40	0.30	0.48	0.47	0.43
READ-POE	0.32	0.45	0.47	0.40	0.33	0.50	0.50	0.45
READ-BIO	0.31	0.43	0.45	0.38	0.31	0.49	0.50	0.45
READ-HST	0.29	0.40	0.40	0.36	0.29	0.44	0.45	0.42
ARITH	0.49	0.55	0.54	0.54	0.36	0.56	0.58	0.53
ALGEBRA	0.46	0.51	0.52	0.52	0.34	0.54	0.56	0.51
GEOMETRY	0.19	0.22	0.20	0.23	0.14	0.23	0.23	0.22
PROBILTY	1.00	0.35	0.33	0.34	0.22	0.35	0.37	0.33
EARTHSCI	0.35	1.00	0.50	0.47	0.33	0.54	0.51	0.49
LIFE SCI	0.33	0.50	1.00	0.43	0.33	0.49	0.49	0.46
CHEMISTR	0.34	0.47	0.43	1.00	0.29	0.45	0.44	0.43
SCI METH	0.22	0.33	0.33	0.29	1.00	0.34	0.34	0.32
HISTORY	0.35	0.54	0.49	0.45	0.34	1.00	0.64	0.55
CIT/GOVT	0.37	0.51	0.49	0.44	0.34	0.64	1.00	0.54
GEOG/EC	0.33	0.49	0.46	0.43	0.32	0.55	0.54	1.00

Source: U.S. Department of Education, National Center for Education Statistics, "National Education Longitudinal Study of 1988: Base Year Survey."

APPENDIX G

DEFINITIONS OF PROFICIENCY SCORES

APPENDIX G

Definitions of Proficiency Scores

Each proficiency score level was marked by four items, which were chosen as having similar difficulty and content. Success, or "passing" a level, was defined as answering at least three of the four items correctly. As described in the text of the report, two such levels were defined for Reading, and three for Mathematics. The sequence numbers of the items selected for determining the proficiency levels are listed below, along with their content classifications and a brief description of the item itself.

Reading

Level 1: Simple reading comprehension including reproduction of detail and/or the author's main thought

1	Repro-Detail	Identify the objective of a character's action
2	Repro-Detail	Identify character's assumption in planning action
3	Repro-Detail	Identify the reason the character's plan didn't work
16	Repro-Detail	Define the meaning of a phrase

Level 2: Ability to make inferences beyond the author's main thought and/or understand and evaluate relatively abstract concepts.

5	Inference/Eval	Choose adage that best fits the lesson to be learned
10	Inference/Eval	Infer the meaning of a metaphor from context of line
13	Inference/Eval	Identify the author's state of mind
14	Inference/Eval	Identify an example of personification

Mathematics

Level 1: Simple arithmetical operations on whole numbers

16	Proc/Decl	Compare two quantities of money expressed differently
17	Proc/Decl	Compare two simple arithmetic expressions involving division of integers
19	Proc/Decl	Compare two simple arithmetic expressions involving multiplication of integers
20	Proc/Decl	Set up a simple equation involving addition or subtraction of integers that is the solution of a word problem

Level 2: Simple operations with decimals, fractions, and roots

5	Proc/Decl	Perform an arithmetic operation (square root) and compare result with a number

13	Proc/Decl	Compare an integer with an expression using division of decimals
14	Proc/Decl	Compare expressions, given information containing exponents
18	Proc/Decl	Compare two simple arithmetic expressions involving division

Level 3: Simple problem solving, requiring conceptual understanding and/or the development of a solution strategy

11	Problem Solving	Compare length of line segments illustrated in a diagram
36	Comprehension	Choose which expression is different from a specified percentage
39	Comprehension	Supply number that completes an algebraic equation correctly
40	Proc/Decl	Simplify an algebraic expression

Assigning students to one of three proficiency categories for Reading (below Level 1, proficient at Level 1 but not Level 2, and proficient at Level 3) and four analogous categories for Mathematics was a straightforward process for the majority of test-takers. Even if a student had omitted one or more items in a 4-item cluster, a pass/fail determination could be made as long as the remaining three items had been answered correctly, or at least two were answered incorrectly.

Problems in identifying a student's proficiency level could arise from one of two conditions. First, a student might not answer enough items at one or more levels to meet either the 3-correct (pass) or 2-incorrect (fail) criterion. This might possibly due to lack of motivation to complete a "no risk" test, or a reluctance to guess that seems to characterizes some students. As pointed out in the text section on speededness, insufficient time to complete the test was unlikely to have been a factor. The second possible problematic response pattern is a "reversal", that is, passing a more difficult level after failing an easier one. Such a reversal pattern might be a result of a few careless mistakes combined with a few lucky guesses, or, again, could be related to motivation. In any case, it would be inconsistent with the hypothesized hierarchical model.

Proficiency scores on the Reading test could be determined directly for 96% of the students who had taken the test. Only about 3% of the students answered too few items to be classified, and 1% had the only possible reversal pattern: fail Level 1, pass Level 2. Success in classifying students on the Reading test was probably due to several factors. The Reading test was the first test in the booklet, so unmotivated students may not yet have gotten tired of responding. Only two levels, eight items, were required, most of which fell in the first part of the test. And with only one reversal pattern possible, the potential for inconsistencies due to guessing was minimal. NCES staff members decided that the 4% rate of unclassified students did not warrant attempts at resolution.

Assignment of Mathematics proficiency scores was a considerably more complex process. Determinations based on the students' item responses alone resulted in only 86% of the students being classified. About 8.5% of the students had omitted too many items to be categorized, and another 5.5% had reversals. Again, several factors were at work. Three of the four Level 3 items fell at or near the end of the Mathematics section, where they were least likely to be answered either by the few students who ran out of time or by those not motivated to finish. Mathematics had more proficiency levels, three, consisting of more items, twelve, than were required for classification in Reading. And the potential for reversals was greater: with three levels, there are four different ways a reversal could occur. The 14% missing data rate for mathematics proficiency scores was unacceptably high. In particular, it appeared that population estimates of mathematics proficiency might be biased upward if a substantial number of the lowest-ability students, who were more likely to have omitted some of the Level 3 items, were not scored. Evidence for this view was provided by the IRT formula score mean for students excluded for missing responses: it was nearly half a standard deviation lower than that of the total sample.

A classification scheme was devised by a consensus of NCES staff and project staff that provided estimates of proficiency levels for about half of the missing Mathematics students.

First of all, it was decided not to attempt resolution of the 5.5% of students who demonstrated reversal patterns. These students did have enough items answered to be scored, but their classifications, for whatever reason, did not fit the hierarchical model. Moreover, since their IRT formula score mean was almost identical to that of the total sample, it appeared that omitting proficiency scores for these students would not introduce any systematic bias into the national estimates.

The procedure for obtaining proficiency scores for students who had omitted critical items required a method of guessing of what those item responses would have been had they been there. The Item Response Theory (IRT) parameters described in the text of the report provided a means of obtaining estimates of item responses for each individual student. The formula presented in that section specifies the probability that a student at a particular ability level, theta, will answer correctly on a specific item, given the three parameters of that item: **a** (discrimination index), **b** (difficulty level), and **c** (the guessing parameter).

A "simulated" right/wrong response to the item can then be obtained by, essentially, flipping a biased coin, with the amount of bias in the coin toss equal to the probability of a correct answer. Translated into operational terms, this means obtaining a computer-generated random number between 0 and 1, and comparing it with the probability of a correct answer provided by the formula. If the random number is less than or equal to the probability, the simulated response is "correct"; otherwise it is "incorrect." For example, if a particular student has a probability of getting a particular item correct equal to .75, then any random number up to and including .75 will produce an estimated correct response; a random number greater than .75 will be classified as incorrect.

Given a procedure for simulating answers to omitted items, NCES staff members specified a set of decision rules for resolutions that took into account the number and location of the missing items. Response patterns were grouped, and treated as described below.

1) All students who omitted items at Level 1, but passed Levels 2 and 3, (designated _PP) were judged to have passed all three levels without resorting to simulation scores for the missing items. It was reasoned that if at least three out of four of the more difficult items were answered correctly at both of the advanced levels, the student almost certainly was proficient at the lowest level as well. Similarly, students who failed the first two levels and omitted Level 3 items (FF_) were assigned a failing score at the highest level. If these students answered sufficient items at the two lower levels, and answered them incorrectly, it was highly unlikely that they possessed the skills to solve three out of four items in the most difficult cluster.

2) The next three patterns treated consisted of students who had answered sufficient items to be classified at two of the three levels, and omitted items only at one level. In addition the location of the missing level, and the right/wrong designation of the remaining two, was such that the missing level could be resolved either way, pass or fail, and still produce a consistent (hierarchical) result. These three patterns were:

   PP_ (Pass Levels 1 and 2, omit items at Level 3)
   P_F (Pass Level 1, omit items at Level 2, fail Level 3)
   _FF (Omit items at Level 1, fail Levels 2 and 3)

   As can be seen, either a P or an F inserted in the blank spaces would produce an acceptable solution. For all students with these three response patterns, item responses were simulated for all omitted items in the blank level, regardless of how many of the four items were blank. Then the simulated correct responses were counted along with the actual correct responses, and a pass/fail score for the missing level was assigned based on the three out of four requirement.

3) The remaining students had response patterns with either a missing designation at more than one level, and/or a pattern that indicated a potential for a reversal. Given the ambiguity, it was decided to implement the simulation procedure for a given level only if two or more items had been responded to at that level. If this relatively conservative treatment yielded either a consistent (hierarchical) pattern, or the _PP or FF_ patterns described in (1.) above, proficiency scores were assigned accordingly. If the constraint on the number of items simulated still left a blank level other than the two specified, or if the resolution produced a reversal pattern, proficiency scores were omitted for the student.

   The resolution process brought the proportion of students with missing proficiency scores down from 14% to 7.3%. Moreover, it brought the discrepancy

in formula score mean for the unscored cases down from half a standard deviation to about a tenth of a standard deviation. This is a good indication that the bias in estimates due to missing data has been considerably reduced.

APPENDIX H

STANDARD ERRORS OF MEASUREMENT AT THETA SCALE POINTS

Appendix H

Standard Errors of Measurement at Theta Scale Points

Theta	Reading	Math	Science	HCG
-3.0000	1.7458	1.4380	1.6365	1.5644
-2.9000	1.6657	1.3598	1.5185	1.3409
-2.8000	1.5881	1.2871	1.4098	1.1543
-2.7000	1.5132	1.2192	1.3102	1.0003
-2.6000	1.4419	1.1555	1.2189	0.8743
-2.5000	1.3741	1.0956	1.1351	0.7719
-2.4000	1.3098	1.03g9	1.0584	0.6895
-2.3000	1.2483	0.9849	0.9883	0.6236
-2.2000	1.1892	0.9331	0.9242	0.5617
-2.1000	1.1313	0.8832	0.8660	0.5314
-2.0000	1.0740	0.8349	0.8132	0.5008
-1.9000	1.0162	0.7880	0.7656	0.4780
-1.8000	0.9575	0.7424	0.7229	0.4617
-1.7000	0.8978	0.6981	0.6850	0.4503
-1.6000	0.8376	0.6552	0.6517	0.4427
-1.5000	0.7778	0.6138	0.6228	0.4377
-1.4000	0.7199	0.5742	0.5980	0.4345
-1.3000	0.6651	0.5365	0.5772	0.4323
-1.2000	0.6147	0.5008	0.5600	0.4304
-1.1000	0.5693	0.4672	0.5460	0.4282
-1.0000	0.5293	0.4358	0.5347	0.4253
-0.9000	0.4946	0.4066	0.5254	0.4215
-0.8000	0.4648	0.3795	0.5171	0.4167
-0.7000	0.4393	0.3547	0.5089	0.4112
-0.6000	0.4175	0.3321	0.4996	0.4050
-0.5000	0.3986	0.3119	0.4884	0.3978
-0.4000	0.3821	0.2939	0.4750	0.3894
-0.3000	0.3674	0.2783	0.4596	0.3792
-0.2000	0.3542	0.2647	0.4429	0.3674
-0.1000	0.3424	0.2530	0.4262	0.3543
0.0000	0.3322	0.2429	0.4105	0.3411
0.1000	0.3241	0.2344	0.3967	0.3291
0.2000	0.3183	0.2273	0.3852	0.3192
0.3000	0.3154	0.2218	0.3759	0.3119
0.4000	0.3157	0.2181	0.3686	0.3071
0.5000	0.3195	0.2163	0.3628	0.3043
0.6000	0.3270	0.2167	0.3583	0.3032
0.7000	0.3381	0.2194	0.3549	0.3035
0.8000	0.3531	0.2247	0.3526	0.3052
0.9000	0.3719	0.2323	0.3517	0.3083
1.0000	0.3948	0.2425	0.3524	0.3128

Appendix H (con'd)

Standard Errors of Measurement at Theta Scale Points (Continued)

Theta	Reading	Math	Science	HCG
1.1000	0.4217	0.2552	0.3551	0.3181
1.2000	0.4528	0.2704	0.3602	0.3240
1.3000	0.4883	0.2883	0.3680	0.3302
1.4000	0.5281	0.3089	0.3788	0.3376
1.5000	0.5725	0.3321	0.3928	0.3475
1.6000	0.6216	0.3581	0.4099	0.3619
1.7000	0.6755	0.3869	0.4102	0.3826
1.8000	0.7343	0.4184	0.4535	0.4107
1.9000	0.7983	0.4528	0.4797	0.4470
2.0000	0.8675	0.4902	0.5084	0.4919
2.1000	0.9420	0.5307	0.5397	0.5454
2.2000	1.0220	0.5745	0.5733	0.6075
2.3000	1.1076	0.6217	0.6094	0.6780
2.4000	1.1987	0.6725	0.6480	0.7569
2.5000	1.2954	0.7272	0.6891	0.8442
2.6000	1.3978	0.7860	0.7328	0.9400
2.7000	1.5055	0.8490	0.7793	1.0445
2.8000	1.6188	0.9165	0.8289	1.1581
2.9000	1.7371	0.9886	0.8814	1.2811
3.0000	1.8605	1.0656	0.9373	1.4139

•U.S. GOVERNMENT PRINTING OFFICE: 1991 • 281 • 697 • 43352

United States
Department of Education
Washington, D.C. 20208–5574

Official Business
Penalty for Private Use, $300

U.S.MAIL

POSTAGE AND FEES PAID
U.S. DEPARTMENT OF EDUCATION
ED 395

FOURTH CLASS RATE

BEST COPY AVAILABLE